



UNITED STATES DEPARTMENT OF COMMERCE
Bureau of the Census
Washington, DC 20233-0001

February 28, 2001

DSSD CENSUS 2000 PROCEDURES AND OPERATIONS MEMORANDUM SERIES B-6*

MEMORANDUM FOR Howard Hogan
 Chief, Decennial Statistical Studies Division

From: David C. Whitford *DCW*
 Assistant Division Chief for Statistical Program Management
 Decennial Statistical Studies Division

Prepared by: Danny R. Childers, Rosemary L. Byrne, Tamara S. Adams, and
 Roxanne Feldpausch
 Decennial Statistical Studies Division

Subject: Accuracy and Coverage Evaluation: Person Matching and Follow-
 up Results

The attached document was prepared, per your request, to assist the Executive Steering Committee on A.C.E. Policy in assessing the data with and without statistical correction.

This report focuses on the person matching and follow-up results.

Accuracy and Coverage Evaluation: Person Matching and Follow- up Results

Danny R. Childers, Rosemary L. Byrne, Tamara S. Adams, and Roxanne
Feldpausch

U.S. Census Bureau

Table of Contents

Executive Summary	1
Introduction	3
Person Matching Results	4
Before follow-up results	5
What are the preliminary census day interview outcome codes?	8
What are nonmatch rates for the P Sample and E Sample people?	14
The Follow-up Interview	25
After Follow-up Match Results	27
What are the P sample not matched and E Sample erroneous enumeration rates after follow-up?	31
What are the types of erroneous enumerations?	40
What are the unresolved rates?	52
What are the final census day interview outcome codes?	59
What are census day noninterview rates for occupied housing units?	62
How do the preliminary and final census day interview outcome codes compare?	66
What are the final interview day interview outcome codes?	68
Quality Assurance of the Clerical Person Matching Operation	73
How was quality assurance operationalized?	73
How was the clerical matching performed?	73
What were the quality assurance operations?	74
What are the assumptions of the QA Plan?	75
What are the limitations on matching QA?	76
What are the QA Results of the clerical matching operation?	76
What are the individual user change rates for technicians in sampled clusters/batches?	76
What are the individual change rates for clerks in sampled clusters/batches? ..	76
How do we estimate the estimated overall change rates for BFU?	77
Conclusions	80
Person Followup Quality Assurance	80
How was person followup quality assurance operationalized?	80
What are the results from person followup QA?	81
Conclusions	83
References	84
Appendix 1—Percentage of the Weighted E Sample for Each Variable	A-1
Appendix 2—Equations for Matching Quality Assurance Calculations	B-1

Tables

Table 1: National P Sample Before Follow-up Matching	7
Table 2: National E Sample Before Follow-up Matching	7
Table 3a: Preliminary Census Day Interviewing Outcome for A.C.E. Housing Units	9
Table 3b: Preliminary Census Day Interviewing Outcome for P Sample Housing Units	10
Table 4a: P sample Preliminary Percent Noninterview Before Follow-up in Occupied Housing Units by Census Regional Office (Unweighted Data)	11
Table 4b: P sample Preliminary Percent Noninterview Before Follow-up in Occupied Housing Units by Census Region (Unweighted Data)	12
Table 4c: P sample Preliminary Percent Noninterview Before Follow-up in Occupied Housing Units by Type of Enumeration Area (Unweighted Data)	13
Table 4d: P sample Preliminary Percent Noninterview in Before Follow-up by Respondent Type (Unweighted Data)	13
Table 4e: P sample Preliminary Percent Noninterview Before Follow-up by Interview Mode (Unweighted Data)	14
Table 5a: Percent Not Matched Before Follow-up by Census Region (Unweighted Data)	15
Table 5b: Percent Not Matched Before Follow-up by Census Regional Office (Unweighted Data)	16
Table 5c: Percent Not Matched Before Follow-up by Type of Enumeration Area (Unweighted Data)	17
Table 5d: Percent Not Matched Before Follow-up by Sex (Unweighted Data)	17
Table 5e: Percent Not Matched Before Follow-up by Age (Unweighted Data)	18
Table 5f: Percent Not Matched Before Follow-up by Race (Unweighted Data)	19
Table 5g: Percent Not Matched Before Follow-up by Hispanic Origin (Unweighted Data) ...	20
Table 5h: Percent Not Matched Before Follow-up by Tenure (Unweighted Data)	20

Table 5i: Percent Not Matched Before Follow-up by Percent Mobile Home (Unweighted Data)	21
Table 5j: Percent Not Matched Before Follow-up by Percent Multi-Unit (Unweighted Data)	21
Table 5k: Percent Not Matched in the P sample Before Follow-up by Type of Address (Unweighted Data)	22
Table 5l: Percent Not Matched in the P sample Before Follow-up by Type of Unit at Listing (Unweighted Data)	23
Table 5m: Percent Not Matched in the P sample Before Follow-up by Mover Status (Unweighted Data)	24
Table 5n: Percent Not Matched in the P sample Before Follow-up by Interview Respondent (Unweighted Data)	24
Table 5o: Percent Not Matched in the P sample Before Follow-up by Interview Mode (Unweighted Data)	25
Table 6a: Results of P sample Follow-up for People Followed up	26
Table 6b: Results of E Sample Follow-up for Nonmatches and Possible Matches	27
Table 7a: National P Sample Match Status After Follow-up	29
Table 7b: National P Sample Residence Status After Follow-up	29
Table 7c: National E Sample Matching After Follow-up	30
Table 7d: Match Status for P Sample After Follow-up	30
Table 8: 1990 Percent Net Undercount from Dual System Estimation by Race and Ethnic Origin	31
Table 9a: Percent Not Matched and Erroneously Enumerated After Follow-up by Census Region Before Weighting and Imputation	32
Table 9b: Percent Not Matched and Erroneously Enumerated After Follow-up by Census Regional Office Before Weighting and Imputation	33
Table 9c: Percent Not Matched and Erroneously Enumerated After Follow-up by Tenure Before Weighting and Imputation	34

Table 9d: Percent Not Matched and Erroneously Enumerated After Follow-up by Type of Enumeration Area Before Weighting and Imputation	34
Table 9e: Percent Not Matched and Erroneously Enumerated After Follow-up by Age and Sex Before Weighting and Imputation	35
Table 9f: Percent Not Matched and Erroneously Enumerated After Follow-up by Simulated Race Domain Before Weighting and Imputation	36
Table 9g: Percent Not Matched and Erroneously Enumerated After Follow-up by Mail Return Rate Before Weighting and Imputation	37
Table 9h: Percent Not Matched and Erroneously Enumerated After Follow-up by Percent Mobile Home Before Weighting and Imputation	38
Table 9i: Percent Not Matched and Erroneously Enumerated After Follow-up by Percent Multi-Unit Before Weighting and Imputation	39
Table 9j: Percent Not Matched After Follow-up by A.C.E. Mover Status Before Weighting and Imputation	39
Table 9k: Percent Not Matched After Follow-up by A.C.E. Interview Mode Before Weighting and Imputation	40
Table 9l: Percent Not Matched After Follow-up by A.C.E. Respondent Type Before Weighting and Imputation	40
Table 10: 1990 Erroneous Enumerations Final Weighted Numbers	41
Table 11a: 2000 Unweighted Types of Erroneous Enumerations	42
Table 11b: Percent of E Sample for Type of Erroneous Enumeration After Follow-up Before Weighting and Imputation by Simulated Race Domain	43
Table 11c: Percent of E Sample for Type of Erroneous Enumeration After Follow-up Before Weighting and Imputation by Type of Enumeration Area	44
Table 11d: Percent of E Sample for Type of Erroneous Enumeration After Follow-up Before Weighting and Imputation by Age and Sex	45
Table 11e: Percent of E Sample for Type of Erroneous Enumeration After Follow-up Before Weighting and Imputation by Tenure	46
Table 11f: Percent of E Sample for Type of Erroneous Enumeration After Follow-up Before Weighting and Imputation by Region	46

Table 11g: Percent of E Sample for Type of Erroneous Enumeration After Follow-up Before Weighting and Imputation by Regional Office	47
Table 11h: Percent of E Sample for Type of Erroneous Enumeration After Follow-up Before Weighting and Imputation by Percent Mobile Homes in the Cluster	48
Table 11i: Percent of E Sample for Type of Erroneous Enumeration After Follow-up Before Weighting and Imputation by Percent Multi-Unit in the Cluster	48
Table 11j: Percent of E Sample for Type of Erroneous Enumeration After Follow-up Before Weighting and Imputation for Simulated Post-Strata Variables	49
Table 12a: Percent Unresolved After Follow-up by Census Region Before Weighting and Imputation	52
Table 12b: Percent Unresolved After Follow-up by Census Regional Office Before Weighting and Imputation	53
Table 12c: Percent Unresolved After Follow-up by Tenure Before Weighting and Imputation	53
Table 12d: Percent Unresolved After Follow-up by Type of Enumeration Area Before Weighting and Imputation	54
Table 12e: Percent Unresolved After Follow-up by Age and Sex Before Weighting and Imputation	55
Table 12f: Percent Unresolved After Follow-up by Simulated Race Domain Before Weighting and Imputation	56
Table 12g: Percent Unresolved After Follow-up by Mail Return Rate Before Weighting and Imputation	57
Table 12h: Percent Unresolved After Follow-up by Percent Mobile Home Before Weighting and Imputation	57
Table 12i: Percent Unresolved After Follow-up by Percent Multi-Unit Before Weighting and Imputation	58
Table 12j: Percent Unresolved After Follow-up by A.C.E. Mover Status Before Weighting and Imputation	58
Table 12k: Percent Unresolved After Follow-up by A.C.E. Interview Mode Before Weighting and Imputation	59

Table 19b: P sample Noninterview Rates for Interview Day in Occupied Housing Units by Census Regional Office (Unweighted Data)	70
Table 19c: P sample Noninterview Rates for Interview Day in Occupied Housing Units by Census Region (Unweighted Data)	70
Table 19d: P sample Noninterview Rates for Interview Day in Occupied Housing Units by Type of Enumeration Area (Unweighted Data)	71
Table 19e: P sample Noninterview Rates for Interview Day in Occupied Housing Units by Whether the Interview was done in the Nonresponse Conversion Operation (NRCO) (Unweighted Data)	71
Table 19f: P sample Noninterview Rates for Interview Day in Occupied Housing Units by Type of Structure from A.C.E. Listing (Unweighted Data)	72
Table 19g: P sample Noninterview Rates for Interview Day in Occupied Housing Units by Status of A.C.E. Listing (Unweighted Data)	72
Table 19h: P sample Noninterview Rates for Interview Day in Occupied Housing Units by Type of Interview (Unweighted Data)	73
Table 20: Coding Changes in Sampled Clusters/Batches by Level of Highest Review	79
Table 21: Overall Change Rate and Outgoing Quality Rate by Stage	79
Table 22: Records Completed by Stage	80
Table 23: Cluster/Batch to User Ratio	80
Table 24: QA Selection Status of PFU Cases	82
Table 25: Workload for PFU QA by A.C.E. Regional Office	82
Table 26: QA Outcome of Cases in PFU QA	83
Table A-1: Percent of the Weighted E Sample by Census Region	A-1
Table A-2: Percent of the Weighted E Sample by Census Regional Office	A-2
Table A-3: Percent of the Weighted E Sample by Tenure	A-2
Table A-4: Percent of the Weighted E Sample by Type of Enumeration Area	A-3
Table A-5: Percent of the Weighted E Sample by Age and Sex	A-4

Table A-6: Percent of the Weighted E Sample by Race Domain	A-5
Table A-7: Percent of the Weighted E Sample by Mail Return Rate	A-5
Table A-8: Percent of the Weighted E Sample by Percent Mobile Home	A-6
Table A-9: Percent of the Weighted E Sample by Percent Multi-Unit	A-6

Accuracy and Coverage Evaluation 2000: Person Matching and Follow-up Results

prepared by Danny R. Childers, Rosemary L. Byrne, Tamara S. Adams, and Roxanne Feldpausch

Executive Summary

We assessed the data's quality in the matching and follow-up for the Accuracy and Coverage Evaluation (A.C.E.). We also assessed the quality of the person interview by looking at noninterview rates for the A.C.E. person interview.

How does the matching in 2000 compare to 1990?

Matching refers to determining whether an individual enumerated in the A.C.E. was the same person as an individual enumerated in the census. Because errors in matching could significantly affect the undercount estimates, highly accurate matching was an important part of the A.C.E. methodology. Neither Secretary Mosbacher nor the Committee on Adjustment of Postcensal Estimates (CAPE) identified matching error as a significant problem with the 1990 PES. Still, the Census Bureau significantly improved matching in the 2000 A.C.E. design, and matching error was expected to be even lower in Census 2000 than in 1990.

What was the quality of the A.C.E. interviewing?

The noninterview rates gave an indication of the quality of the interviewing for the A.C.E. Based on this indication, the quality of the 2000 interview exceeded the quality of the 1990 interview. In 1990, noninterview rate for the current residents was 1.6 percent. In 2000, the interview day percent noninterview was 1.1 percent which exceeded our goal of having a noninterview rate below 2.0 percent.

We expected the noninterview rate for census day residents to be greater than the noninterview rate for interview day. In 2000, the noninterview rate for the census day residents was 3.0 percent which was smaller than we expected indicating quality in our A.C.E. interviewing. This noninterview rate includes noninterviews resulting from incomplete interviews. A noninterview rate of 3.0 percent for census day residents means 97.0 percent of the occupied census day households were interviewed.

What was the quality of the clerical matching?

We use change rates as an indication of the quality of the clerical matching. We first calculated individual change rates of significant code changes for each clerk and technician based on their clusters sampled for QA review. We assumed matching analysts had no error. We then estimated overall change rates for the records worked in clerical matching using three different models.

We did not consider the computer matched records. For all three models, the change rates were as follows:

- Before Followup Clerks ranged from 0.44 percent change to 0.59 percent change
- Before Followup Technicians ranged from 0.20 percent to 0.23 percent change
- After Followup Clerks ranged from 0.11 percent change to 0.95 percent change
- After Followup Technicians ranged from 0.13 percent change to 0.71 percent change

Matching QA was successful at minimizing errors. The outgoing quality rates are higher than 99% for all levels in both stages.

What are the results from person followup quality assurance?

Of 89,334 cases sent to person followup, quality assurance considered 8,929 cases (4,491 sampled, 4,438 targeted). We considered 0.45 percent of the randomly sampled cases and 0.94 percent of the targeted cases possibly discrepant. We considered the rest of cases not part of the quality assurance random sample for person followup (84,843) to have a rate of discrepancy similar to that of the randomly sampled cases. In addition, we corrected 84 of those cases in the targeted sample.

Introduction

This memorandum documents the results of the person matching and follow-up operations. The person matching results are presented for both before and after follow-up matching. We also documented the noninterview rates for the A.C.E. person interview. Quality assurance results are presented for each step in the clerical matching and follow-up interviewing operations.

Several deficiencies in the 1990 Post-Enumeration Survey (PES) design and matching operations prompted improvements in the 2000 A.C.E. matching operation. One problem in 1990 was the misreporting of census day address, with an estimated 0.7 percent of the P sample being erroneously reported as nonmovers (West 1991). The 2000 A.C.E. improved on 1990 PES in several ways.

- The Computer Assisted Personal Interview (CAPI) instrument improved the quality of the reporting of mover status because it was a more automated process.
- In 1990, each inmover household (those that moved into PES block clusters after census day) had to be matched to a census day address, which was usually outside the cluster. In 2000, the reconstructed census day household was matched to the census enumerations in the sample block cluster.

The census day household consists of nonmovers and outmovers. The nonmovers lived in the housing unit at the time of the interview and on census day. The outmovers lived in the housing unit on census day, but moved before the A.C.E. interview. Nonmovers and outmovers were matched to people in their block cluster. For clusters with high rates of A.C.E. housing unit nonmatch and census geocoding error, the search area was extended in Census 2000 to the surrounding blocks. Note that the unresolved match codes due to incomplete mover address did not exist in A.C.E. 2000 because there was no inmover matching operation.

A study of clerical error in the 1990 PES found error in coding matches (Davis 1991a) and erroneous enumerations (Davis 1991b). In 1990, codes were entered into a computer system, but the actual matching and duplicate searches were done using paper. In 2000 A.C.E., the matching was better controlled and more efficient than 1990 because the clerical matching and quality assurance were automated instead of on paper and coded into a system. The automated interactive system did not prevent all matching error, but reduced the chances for error significantly. Examples of the improvements in coding include:

- Electronic filtering allowed searching based on first name, last name, characteristics, and addresses. For example, the system allowed searching for all people named George, all people whose last name begins with an H, all people on Elm Street, or everyone in the 30 to 40 age range.
- Only particular codes that fit the situation were allowed. For example, only P sample nonmatch codes can be assigned a P sample nonmatch after follow-up code.

- The electronic searches for duplicates reduced the tedious searching through paper lists of census people. The searching in 1990 was limited to printouts in two sorts: last name and household by address. In 2000, the clerks had the capability to filter on name, characteristics, and address to help identify duplicates.
- The system monitored whether the matcher had completed all the necessary searches such as looking for duplicates.
- There were built-in edits to check for consistent coding. For example, codes that apply to a household were assigned to all people in the household, such as a geographic code.
- The system automatically assigned certain codes, minimizing coding error.
- Clerical matchers could use a code indicating the case needs review at the next level of matching. This code allowed them to flag unusual cases to be done by a person with more experience.
- All quality assurance for the clerical matching was automated.
- Clerical matching was centralized at the National Processing Center instead of different groups of matchers in the seven processing offices, as was done in 1990. Forty-six Technicians were hired in September 1999 and were thoroughly trained in the design of the A.C.E. and matching of people and housing units. These Technicians performed the quality assurance for the clerical matchers. Additionally, ten Analysts were the most experienced matchers. The Analysts did the quality assurance for the Technicians and handled the most difficult cases.
- We also ran computer programs on the completed before follow-up matching results to identify clusters where the programs detected matches and duplicates not identified by the clerical matchers. Consistency checks were also performed between housing unit and person match codes.
- Keying error in the data capture of the 1990 PES was reduced because the 2000 interview used a CAPI instrument. A more accurate capture of the data increased the efficiency of the computer matching.

Person Matching Results

The P sample people that were eligible for matching are the nonmovers, the outmovers, and the people with unresolved residence status from the A.C.E. interviewing. These census day residents should have been enumerated in the census. The in-movers were not included in the P sample for matching. The E sample included the corresponding census people in the sample block cluster. Census people not in sample after subsampling of housing units in large block clusters and people in group quarters were not included in the E sample.

The P sample and census people within the sample block cluster were computer matched followed by a clerical review. The matching steps are:

- First, we matched between the P sample and the E sample.
- The remaining not matched P sample people were clerically searched among the non-E Sample people in the sample block cluster, which includes the people enumerated in

- group quarters and census people in the block cluster subsampled out.
- P sample duplicates and E sample duplicates were identified clerically.
- In clusters selected for targeted extended search, matchers searched the surrounding blocks clerically for P sample matches and possible matches. Census people geocoded in the surrounding blocks were coded as correctly enumerated.

P-sample and E sample nonmatches were sent for a follow-up interview. The results of the interview were clerically recorded in the matching software. The results of before and after follow-up coding were displayed to monitor the data in order to identify anomalies in the A.C.E. and census data. Note that the purpose of this analysis is to point out important features of the matching and quality assurance procedures. No formal tests of statistical hypotheses have been performed. These data are unweighted and were tabulated from the person matching files.

Before follow-up results

The first two tables contain the results of before follow-up matching for the P sample and the E sample. For details of these codes, see Childers (2000). These before follow-up matching results are from unweighted data from the fifty states and the District of Columbia. These tables do not include the before follow-up matching results in Puerto Rico. The P sample codes are grouped into

- Matched
- Not matched
- Possible match
- Unresolved match status
- Removed from the P sample

Matched - The P sample person was found in the census.

Not Matched - The P sample person was not found in the census. A follow-up interview was conducted for:

- partial household nonmatches
- whole households of conflicting household members (i.e., whole households of P sample and census nonmatches)¹
- other whole household nonmatches where the P sample interview was conducted with a nonhousehold member²

¹ These cases have been called the Smith/Jones cases in the past.

² No follow-up interview was conducted when there were whole households of P sample nonmatches from interviews with household members in a housing unit that did not match in the housing unit operation or matched to a housing unit containing no data-defined people.

Possible Match - The P sample person may have been a match to the census person. A follow-up interview was needed to determine if the two names referred to the same person.

Unresolved Match Status - The only category of unresolved before follow-up was insufficient information for matching and follow-up for the P sample person.

Removed from the P sample - The only category of removed from the P sample in the before follow-up matching were the P sample people coded as duplicates. The P sample duplicates are removed because they were listed more than once.

The E sample codes are grouped into

- Correctly enumerated
- Erroneously enumerated
- Not matched and needing a follow-up interview
- Possible match
- Unresolved

Correctly enumerated - The only correctly enumerated people in before follow-up matching were the ones matching the P sample.

Erroneously enumerated - The categories during before follow-up were fictitious people, duplicates, insufficient information for matching and follow-up, and geocoding errors.

- The fictitious people were ones where we found notes on the census image identifying the person as not a real person such as a dog or other pet.
- The E sample people enumerated more than once were coded as duplicates.
- The E sample people with insufficient information for matching and follow-up were those who were data-defined, but did not contain full name and at least two characteristics.³
- Census people in housing units identified as geocoding errors⁴ during the initial housing unit follow-up were coded as erroneously enumerated because of geocoding error.

E Sample nonmatches - All E sample people who did not match to the P sample were sent for a follow-up interview.

³ This is the same rule that was used in the 1990 PES. There must have been enough information about the person to have a chance at locating the person for a follow-up interview before the person was allowed into the matching process. See Childers (2000).

⁴ A geocoding error is an error in assigning the housing unit to the correct location.

E Sample possible matches - E sample people who were coded as possible matches were followed up to determine whether they were, in fact, matches.

Unresolved enumeration status - In before follow-up matching, the unresolved category only includes the census housing units that needed targeted extended search field work that was not done.

Tables 1 and 2 contain the results of before follow-up matching for the P sample and the E sample. The before follow-up matching identifies people to be sent for a follow-up interview. See the section on the follow-up interview for the types of people followed up. The number of people coded unresolved will increase after follow-up because of unsuccessful follow-up interviews.

Table 1: National P Sample Before Follow-up Matching

P Sample Match Status	Unweighted People	Percent
Matched	573,506	85.7
Not Matched	76,804	11.5
Possible Match	5,070	0.8
Unresolved	7,524	1.1
Removed	5,923	0.9
Total	668,827	100.0

Table 2: National E Sample Before Follow-up Matching

E Sample Enumeration Status	Unweighted People	Percent
Correctly Enumerated	544,995	76.4
Erroneously Enumerated	27,934	3.9
Not Matched	134,916	18.9
Possible Match	4,751	0.7
Unresolved	304	0.0
Total	712,900	100.0

What are the preliminary census day interview outcome codes?

The preliminary interview outcome codes identified interviews and noninterviews in occupied housing units, vacant housing units, and housing units that are removed from the P sample. The interview outcomes in “Accuracy and Coverage Evaluation Survey: Person Interviewing”, (Byrne 2001) were the interview outcomes for interview day. The interview outcomes described in this section were census day interview outcomes after data editing, which converts whole households of census day residents with insufficient information for matching to noninterviews and whole households of census day residents who should not have been counted at the housing unit on census day to vacant housing units.

Interviews -

- Complete interviews - interviews conducted with a household member.
- Proxy interviews - interviews conducted with someone outside the household
- Sufficient partial interviews - interviews with household members or proxies that collected not all information, but enough information to be considered complete interviews.

Noninterviews -

- Field noninterview
- Whole households of people with insufficient information for matching and follow-up

Vacant on Census Day -

- Housing units identified as vacant on census day by the interviewer
- Whole households of people who should have been counted elsewhere on census day (i.e., whole household nonresidents)

Not a Housing Unit on Census Day -

- The housing units identified during the person interview as not a housing unit on census day are removed from the P sample.

Table 3a contains the number of each category of preliminary outcome code and percentages of each one as a percent of all housing units sent for interview. The interviewers identified 3.4 percent of the A.C.E. addresses as not being housing units on census day.

Table 3a: Preliminary Census Day Interviewing Outcome for A.C.E. Housing Units

Outcome Code	Unweighted Housing Units	Percent
Interview		
Complete interview with a household member	235,632	78.3
Complete interview with a proxy respondent	19,380	6.4
Sufficient partial interview	2,612	0.9
Noninterview		
Field noninterview	2,667	0.9
All people have insufficient information for matching and follow-up	2,321	0.8
Vacant		
No census day residents	4,184	1.4
Vacant on census day	23,911	7.9
Not a Housing Unit		
Not a housing unit on census day	10,206	3.4
Total	300,913	100.0

The A.C.E. housing units identified as not being housing units were not in the P sample. The next table contains the number and percentages of total housing units for the preliminary outcome codes grouped into interview, noninterview, and vacant. The percentages of interview and noninterview for occupied housing units is also included. The noninterview rate for occupied housing units is 1.9 percent based on the preliminary outcome codes before clerical matching.

Table 3b: Preliminary Census Day Interviewing Outcome for P Sample Housing Units

Outcome Code	Total P Sample Housing Units		Occupied P Sample Housing Units	
	Unweighted	Percent	Unweighted	Percent
Interview	257,624	88.6	257,624	98.1
Noninterview	4,988	1.7	4,988	1.9
Vacant	28,095	9.7		
Total	290,707	100.0	262,612	100.0

The percent noninterview was calculated for the unweighted numbers of noninterviews divided by the occupied interviews, which was the interviews plus the noninterviews. Tables of preliminary noninterview rates are presented for several variables. The P Sample preliminary noninterview rates for occupied housing units before the follow-up interview look fairly consistent for Regional Office, census region, and type of enumeration area. The Boston Regional Office had a low noninterview rate because they used additional procedures to convert census day noninterviews.

Table 4a: P Sample Preliminary Percent
Noninterview Before Follow-up in Occupied
Housing Units by Census Regional Office
(Unweighted Data)

Census Regional Office	P Sample Preliminary Percent Noninterview
Boston	0.2
New York	2.6
Philadelphia	2.5
Detroit	1.7
Chicago	1.7
Kansas City	1.5
Seattle	1.8
Charlotte	3.2
Atlanta	2.2
Dallas	2.1
Denver	1.4
Los Angeles	1.9
Total	1.9

Table 4b: P Sample Preliminary Percent
Noninterview Before Follow-up in Occupied
Housing Units by Census Region (Unweighted
Data)

Census Region	P Sample Preliminary Percent Noninterview
Northeast	1.4
Midwest	1.6
South	2.5
West	1.8

The Census Bureau defined type of enumeration area (TEA) codes at the census collection block level. Each block had a TEA code, and no block had more than one TEA code.

- Mailout/Mailback - The USPS delivered the census questionnaires to city-style addresses by mail and the respondent returned the questionnaire by mail or they were visited during nonresponse follow-up.
- Update/Leave - Address lists were compiled by address listing. The questionnaires were delivered by enumerators who updated the address lists when new addresses were discovered. The respondent returned the completed questionnaire by mail or they were visited during nonresponse follow-up.
- List/Enumerate - Enumerators visited these remote and sparsely populated areas simultaneously listing the housing units and enumerating the residents.
- Rural Update/Enumerate - The enumerators began with address lists (for previously update/leave areas), updated their lists of addresses, and completed census questionnaires for the residents.
- Urban Update/Leave - The enumerators began with address lists (for previously mail out/mail back areas), updated their lists of addresses, and left census questionnaires for the residents to complete and return by mail or they were visited during nonresponse follow-up.
- Urban Update/Enumerate - The enumerators began with address lists (for previously mail out/mail back areas), updated their lists of addresses, and completed census questionnaires for the residents.
- Additions to Address Listing Universe of Blocks - Some blocks in the mail out/mail back universe contained a significant number of non-city style address. They were converted to update/leave.

Table 4c: P Sample Preliminary Percent
Noninterview Before Follow-up in Occupied
Housing Units by Type of Enumeration Area
(Unweighted Data)

Type of Enumeration Area	P Sample Preliminary Percent Noninterview
Mail Out / Mail Back	2.0
Update/Leave	1.6
List/Enumerate	0.9
Rural Update/Enumerate	1.2
Urban Update/Leave	1.2
Urban Update/Enumerate	2.3
Additions to Address Listing	1.1

Of all interviews at occupied housing units 7.2 percent were proxy interviews and 92.8 percent were interviews with household members.

Table 4d: P Sample Preliminary Percent
Noninterview in Before Follow-up by
Respondent Type (Unweighted Data)

Respondent Type	P Sample Preliminary Percent Noninterview
Household member	0.9
Proxy	13.8
Total	1.9

Of all interviews at occupied housing units, 33.5 percent were completed by telephone, 66.1 percent were completed by personal visit, and 0.3 percent, which is 910 interviews, were completed by a quality assurance replacement interview. The percent noninterview of occupied housing units for the interview mode is in Table 4e.

Telephone interviews were more likely than personal interviews to have insufficient information because we only have one household to get the information from. There was no opportunity to get better information from a different respondent. Also, there were telephone interviews where we talked to the in-mover and they did not have information about the out-mover. If the people moved into the address after census day, completed the census questionnaire and mailed it back, we could have called an in-mover.

There were several reasons for a high noninterview rate for the quality assurance replacement interviews. These were difficult interviews because they failed the quality assurance check and needed a reinterview. Many of the noninterviews were refusals. Additionally, because the instrument is monitoring both the quality assurance case and the replacement interview, it was difficult to obtain the census day residents in mover cases and many of these were noninterviews. There was also a problem with the instrument in cases where the quality assurance interviewer could not find the address. In these cases, the case failed the quality assurance check but no data was collected for the replacement interview since the address did not exist on the day of the interview. These are considered noninterviews because the cases closed up before any census day information could be obtained. There were 108 of these cases.

Table 4e: P Sample Preliminary Percent
Noninterview Before Follow-up by Interview
Mode (Unweighted Data)

Interview Mode	P Sample Preliminary Percent Noninterview
Telephone	0.9
Personal Visit	2.2
Quality Assurance Replacement	36.0
Total	1.9

What are nonmatch rates for the P sample and E sample people?

The P sample nonmatch rate was calculated by dividing the unweighted number of P sample nonmatches by the unweighted number of P sample people, which was the matches, nonmatches, and unresolved P-sample people. The P sample did not include the people coded as removed. The E sample nonmatch rate was also the unweighted number of E sample nonmatches divided by the unweighted E sample. The percentage not matched for the P sample and E sample is presented next for several variables.

The percent not matched in the Midwest Region in Table 5a appears to be lower for both the P sample and the E sample.

Table 5a: Percent Not Matched Before Follow-up by
Census Region (Unweighted Data)

Census Region	P Sample Percent Not Matched	E Sample Percent Not Matched
Northeast	12.0	19.1
Midwest	8.9	13.2
South	12.6	22.4
West	12.3	19.1
Total	11.6	18.9

In Table 5b the Detroit and Kansas City Regional Offices are in the Midwest Region and appear to be lower than the other Regional Offices.

Table 5b: Percent Not Matched Before Follow-up by Census Regional Office (Unweighted Data)

Census Regional Office	P Sample Percent Not Matched	E Sample Percent Not Matched
Boston	11.8	17.4
New York	13.6	23.3
Philadelphia	12.2	20.6
Detroit	8.2	12.8
Chicago	10.3	14.3
Kansas City	8.4	13.3
Seattle	11.3	18.0
Charlotte	11.9	22.0
Atlanta	12.9	20.8
Dallas	13.1	24.3
Denver	13.4	17.8
Los Angeles	11.4	20.2
Total	11.6	18.9

In Table 5c, the percent not matched is smaller in the mail out/mail back and update/leave areas than the other types of enumeration areas. The mail out/mail back areas are 82.1 percent, the update/leave areas are 16.8 percent, and the other types of enumeration areas are 1.1 percent of the weighted E sample in Table A-4 in Appendix 1.

Table 5c: Percent Not Matched Before Follow-up by Type of Enumeration Area (Unweighted Data)

Type of Enumeration Area	P Sample Percent Not Matched	E Sample Percent Not Matched
Mail Out / Mail Back	11.2	19.1
Update/Leave	11.4	16.8
List/Enumerate	18.3	42.2
Rural Update/Enumerate	20.7	18.4
Urban Update/Leave	10.9	23.2
Urban Update/Enumerate	13.4	13.1
Additions to Address Listing	17.6	21.9
Total	11.6	18.9

There appears to be a slightly higher percent not matched in the P sample and in the E sample for males than females in Table 5d.

Table 5d: Percent Not Matched Before Follow-up by Sex (Unweighted Data)

Sex	P Sample Percent Not Matched	E Sample Percent Not Matched
Male	12.3	19.4
Female	10.8	18.5
Blank	14.5	18.7
Total	11.6	18.9

The age group 18 to 29 appears to have a higher percent not matched for both P sample and E sample in Table 5e.

Table 5e: Percent Not Matched Before Follow-up by Age
(Unweighted Data)

Age	P Sample Percent Not Matched	E Sample Percent Not Matched
Under 18	12.1	18.9
18 to 29	16.5	25.9
30 to 49	10.8	17.1
Over 50	8.7	16.3
Blank	14.7	23.0
Total	11.6	18.9

The percent not matched for the race category of white alone appears to be lower than for the other race classifications for both the P sample and the E sample in Table 5f. The race variable in this table comes from the matching data base.

Table 5f: Percent Not Matched Before Follow-up by Race (Unweighted Data)

Race	P Sample Percent Not Matched	E Sample Percent Not Matched
White alone	9.7	17.3
Black alone	15.8	24.1
American Indian alone	20.9	19.4
Asian alone	12.1	21.0
Native Hawaiian and Pacific Islander alone	16.9	29.6
Other Race alone	14.7	24.3
Multiple Race	13.0	20.4
Blank	18.4	19.6
Total	11.6	18.9

The P sample and E sample people identified as Hispanic appear to have a higher percent not matched than the non-Hispanics in Table 5g.

Table 5g: Percent Not Matched Before Follow-up by
Hispanic Origin (Unweighted Data)

Hispanic Origin	P Sample Percent Not Matched	E Sample Percent Not Matched
Hispanic	14.4	22.6
Non-Hispanic	11.0	18.3
Blank	17.5	19.8
Total	11.6	18.9

The people identified as owners appear to have a lower percent not matched than the renters in Table 5h.

Table 5h: Percent Not Matched Before Follow-up by
Tenure (Unweighted Data)

Tenure	P Sample Percent Not Matched	E Sample Percent Not Matched
Owner	9.2	14.5
Renter	16.1	27.1
Blank	16.0	22.4
Total	11.6	18.9

People in clusters with large amounts of mobile homes appear to have a larger percent not matched for the P sample and the E sample in Table 5i.

Table 5i: Percent Not Matched Before Follow-up by
Percent Mobile Home (Unweighted Data)

Percent Mobile Home	P Sample Percent Not Matched	E Sample Percent Not Matched
None	11.0	18.7
10 Percent or less	11.4	16.7
11 to 50 percent	13.8	19.5
Greater than 50 percent	17.3	26.5
Total	11.6	18.9

People in clusters with large amounts of multi-units appear to have a larger percent not matched for the P sample and the E sample in Table 5j.

Table 5j: Percent Not Matched Before Follow-up by
Percent Multi-Unit (Unweighted Data)

Percent Multi-Unit	P Sample Percent Not Matched	E Sample Percent Not Matched
None	10.7	17.4
10 Percent or less	9.7	16.3
11 to 50 percent	11.0	17.9
Greater than 50 percent	15.0	24.4
Total	11.6	18.9

The next set of tables contains percent not matched in the P sample for variables that are only on the P sample data. People in single units appear to have a smaller percent not matched for the P sample in Table 5k.

Table 5k: Percent Not Matched in the P Sample Before Follow-up by Type of Address (Unweighted Data)

Type of Address	P Sample Percent Not Matched
Single Unit	9.7
Multi-Unit	16.2
Mobile Home not in a Park	16.7
Mobile Home in a Park	20.2
Single Housing Unit in a Special Place	27.2
Multi-Unit in a Special Place	35.7
Other	45.9
Total	11.6

The next table contains the P sample percent not matched for the original classification of the P sample housing unit listed in the Fall of 1999. For example, a P sample housing unit may have been listed as future construction in the Fall of 1999 and existed as a housing unit during the initial housing unit follow-up operation. An interview was conducted at the newly constructed housing unit and 38.6 percent of the P sample census day residents were not found in the census within the search area. People in addresses identified as housing units appear to have a smaller percent not matched for the P sample in Table 51.

Table 51: Percent Not Matched in the P Sample Before Follow-up by Type of Unit at Listing (Unweighted Data)

Type of Unit at Listing	P Sample Percent Not Matched
Housing Unit	11.4
Under Construction	29.7
Future Construction	38.6
Unfit for Habitation	46.3
Boarded Up	29.2
Storage	23.1
Vacant Trailer Site	23.3
Other	70.3
Total	11.6

The P sample people for matching are classified as nonmovers, outmovers, and people with unresolved residence status. The P sample is 95.0 percent nonmovers, 3.5 percent outmovers, and 1.5 percent unresolved residence status. P sample people identified as nonmovers appear to have a lower percent not matched than outmovers and people with unresolved residence in Table 5m.

Table 5m: Percent Not Matched in the P Sample Before Follow-up by Mover Status (Unweighted Data)

Mover Status	P Sample Percent Not Matched
Nonmover	10.9
Outmover	26.6
Unresolved Mover Status	23.6
Total	11.6

The P sample people in housing units with proxy interviews have a larger percent not matched than people obtained from an interview with a household member in Table 5n.

Table 5n: Percent Not Matched in the P Sample Before Follow-up by Interview Respondent (Unweighted Data)

Interview Respondent	P Sample Percent Not Matched
Household Member	10.9
Proxy	24.9
Total	11.6

Telephone interviews were conducted for some households where the census questionnaire was a mail return with a telephone number in Table 5o. The percent not matched in these households was expected to be lower than for other households.

Table 5o: Percent Not Matched in the P Sample Before Follow-up by Interview Mode (Unweighted Data)

Interview Mode	P Sample Percent Not Matched
Telephone	2.1
Personal Visit	16.5
Quality Assurance	16.9
Total	11.6

The Follow-up Interview

The person follow-up is conducted to gather additional information to accurately code the residence status of the nonmatched P sample people and the enumeration status of the E sample people. The following cases were sent to person follow-up:

- P-sample partial household nonmatches
- P-sample whole household nonmatches where the census enumerated different E sample people (i.e., conflicting households or Smith/Jones cases)
- P-sample whole household nonmatches where the A.C.E. person interview was with a proxy respondent
- E-sample nonmatches
- Possible matches between the P sample and the census
- P-sample matches and nonmatches with unresolved residence status

The results of the follow-up interview were clerically entered into the matching software. Table 6a contains the results of the follow-up interviews for the P sample people followed up. The P sample people followed up were classified as

- Matched
- Not matched resident of the cluster on census day
- Unresolved residence or match status
- Nonresident of the cluster on census day and removed from the P sample

Matched - The P sample person was found in the census in the block cluster or in a surrounding block.

Not Matched resident of the cluster on census day - The P sample nonmatch was not found in the census and should have been counted in the search area for this cluster.

Unresolved residence or match status - The person had unresolved residence status because the follow-up interview did not successfully collect the information required to accurately identify this person as a resident of the cluster on census day. In the case of possible matches, the interview did not accurately identify the match status of the people.

Removed from the P sample - The P sample person was not a resident of the housing unit on census day and was removed from the P sample. These people were duplicates, fictitious, living in a P sample housing unit that was listed in the cluster in error (i.e., P sample geocoding error), or the P sample person should have been counted at another residence on census day.

The results of the follow-up interview for the P sample people identified as needing a follow-up interview in Table 6a indicate 14.7 percent unresolved and 12.5 percent removed from the P sample.

Table 6a: Results of P Sample Follow-up for People Followed up		
After Follow-up Match Code	Unweighted People	Percent
Matched	9,793	19.4
Nonmatch Resident	26,961	53.4
Unresolved	7,451	14.7
Removed	6,296	12.5
Total	50,501	100.0

Table 6b contains the results of the follow-up interviews for the E sample people followed up. The E sample people followed up are classified as

- Matched
- Correctly enumerated
- Erroneously enumerated
- Unresolved

Matched - The P sample and E sample people refer to the same person.

Correctly enumerated - The E sample nonmatch was correctly enumerated in the census.

Erroneously enumerated - The E sample nonmatch was erroneously enumerated in the census, because the person should have been counted at another residence on census day, was fictitious, had insufficient information for matching and follow-up, was duplicated, or lived in a household that was a geocoding error.

Unresolved - The follow-up interview for the census nonmatch was not successful.

The results of the E sample follow-up in Table 6b indicate 7.4 percent E sample people followed up were erroneously enumerated and 14.1 percent were unresolved.

Table 6b: Results of E Sample Follow-up for
Nonmatches and Possible Matches

After Follow-up Match Code	Unweighted People	Percent
Matched	9,088	6.3
Correctly Enumerated	103,589	72.2
Erroneously Enumerated	10,618	7.4
Unresolved	20,185	14.1
Total	143,480	100.0

After Follow-up Match Results

The final P sample results are in Tables 7a and 7b. The P sample people have been classified as matched, not matched, unresolved match status, and removed in Table 7a and also tabulated as resident, nonresident, and unresolved residence status in Table 7b. The data are unweighted but the people sampled out of the targeted extended search are removed from tabulations for this section.

The P sample match status is defined as

- matched
- not matched
- unresolved match status
- removed from the P sample

Matched - The P sample was found in the cluster or in the surrounding block in either a housing unit or in group quarters.

Not matched - The P sample person was not found in the search area. If the nonmatch was sent to follow-up, the person was confirmed to be a resident of the cluster on census day. If the nonmatch was not sent for a follow-up interview, a household member identified the person as a resident of the housing unit during the original A.C.E. interview.

Unresolved match status - The match status was unresolved for possible matches with unsuccessful follow-up interviews and for P sample people with insufficient information for matching and follow-up.

Removed from the P sample - People were removed from the P sample when they were fictitious, duplicates, geocoding errors, or not residents of the housing unit on census day.

The P sample residence status was defined as

- resident
- nonresident
- unresolved residence status

Resident - The P sample matched or not matched person was a resident of the housing unit on census day.

Nonresident - P sample people were nonresidents of the cluster when they were fictitious, duplicates, geocoding errors, or should not have been included as a resident of the housing unit on census day. Nonresidents were removed from the P sample.

Unresolved Residence Status - A matched or not matched P sample person had unresolved residence status when the follow-up interview did not successfully determine the person's residence on census day. The residence status of the possible match was unresolved when the follow-up interview was not successful. The residence status was also imputed when the P sample person had insufficient information for matching.

The final E sample results are in Table 7c. The E sample people were classified as correctly or erroneously enumerated and enumeration status of unresolved. These were the unweighted match results that go to imputation and estimation with the people sampled out of the targeted extended search removed.

The E sample enumeration status was defined as

- correctly enumerated
- erroneously enumerated
- unresolved enumeration status

Correctly Enumerated - E sample people were correctly enumerated when they were matched to the P sample or when they have been followed up and they should have been enumerated in this cluster.

Erroneously Enumerated - E sample people were erroneously enumerated when they have another residence where they should have been counted on census day, were fictitious, were duplicated, lived in a housing unit that was a geocoding error, or had insufficient information for matching and follow-up.

Unresolved Enumeration Status - E sample people had unresolved enumeration status when the follow-up interview was unsuccessful. The E sample person may have been followed up to obtain information about the E sample nonmatch, possible match, matched person with unresolved residence status, or geographic work to obtain the location of the housing unit.

Table 7a: National P Sample Match Status After Follow-up

P Sample Match Status	Unweighted People	Percent
Matched	578,695	88.6
Not Matched	54,424	8.3
Unresolved	7,826	1.2
Removed	12,393	1.9
Total	653,338	100.0

Table 7b: National P Sample Residence Status After Follow-up

P Sample Residence Status	Unweighted People	Percent
Resident	625,863	95.8
Nonresident	12,393	1.9
Unresolved	15,082	2.3
Total	653,338	100.0

Table 7c: National E Sample Matching After Follow-up

E Sample Enumeration Status	Unweighted People	Percent
Correctly Enumerated	652,390	92.6
Erroneously Enumerated	31,064	4.4
Unresolved	21,148	3.0
Total	704,602	100.0

The final P sample for matching is in Table 7d after the people are removed from the P sample.

Table 7d: Match Status for P Sample After Follow-up

P Sample Match Status	Unweighted People	Percent
Matched	578,695	90.3
Not Matched	54,424	8.5
Unresolved	7,826	1.2
Total	640,945	100.0

Table 8 contains the net undercount from the PES in 1990 by race and ethnic origin. See Hogan (1993).

Table 8: 1990 Percent Net Undercount from Dual System Estimation by Race and Ethnic Origin

Race and Ethnic Origin	Percent Undercount
Non-Hispanic White and Other	0.7
Black	4.6
Hispanic	5.0
Asian and Pacific Islander	2.4
Reservation Indian	12.2
Total	1.6

What are the P sample not matched and E sample erroneous enumeration rates after follow-up?

The percent P sample not matched and E sample erroneous enumeration is contained in the next set of tables. The percent P sample not matched was one hundred times the nonmatch rate.

$$\text{Nonmatch Rate} = \frac{\text{Not Matched}}{\text{Matched} + \text{Not Matched}}$$

The percent E sample erroneous enumeration was one hundred times the erroneous enumeration rate.

$$\text{Erroneous Enumeration rate} = \frac{\text{Erroneous Enumeration}}{\text{Correct Enumeration} + \text{Erroneous Enumeration}}$$

Both percentages were of unweighted resolved people. The weighting and imputation process happens after the matching was completed. People with a targeted extended search weight of zero, meaning they were sampled out of the targeted extended search, are not included in these tables.

The Midwest Region appears to have a lower P sample percent not matched and E sample percent erroneous enumeration in Table 9a.

Table 9a: Percent Not Matched and Erroneously Enumerated After Follow-up by Census Region Before Weighting and Imputation

Census Region	P Sample Percent Not Matched	E Sample Percent Erroneous Enumeration
Northeast	8.8	5.3
Midwest	6.3	3.6
South	9.5	4.9
West	9.3	4.6
Total	8.6	4.6

The Detroit and Kansas City Regional Offices appear to have a lower P sample percent not matched and E sample percent erroneous enumeration in Table 9b:

Table 9b: Percent Not Matched and Erroneously Enumerated
After Follow-up by Census Regional Office Before
Weighting and Imputation

Census Regional Office	P Sample Percent Not Matched	E Sample Percent Erroneous Enumeration
Boston	8.0	4.6
New York	11.1	7.1
Philadelphia	9.0	4.5
Detroit	5.9	3.4
Chicago	7.6	4.2
Kansas City	5.5	3.5
Seattle	8.4	5.0
Charlotte	8.6	4.2
Atlanta	9.6	5.8
Dallas	10.3	5.0
Denver	9.1	4.4
Los Angeles	9.5	4.0
Total	8.6	4.6

The owners appear to have a lower P sample percent not matched and E sample percent erroneous enumeration than the renters in Table 9c.

Table 9c: Percent Not Matched and Erroneously Enumerated After Follow-up by Tenure Before Weighting and Imputation

Tenure	P Sample Percent Not Matched	E Sample Percent Erroneous Enumeration
Owner	6.2	3.0
Renter	13.2	5.9
Blank	13.5	19.0
Total	8.6	4.6

Table 9d: Percent Not Matched and Erroneously Enumerated After Follow-up by Type of Enumeration Area Before Weighting and Imputation

Type of Enumeration Area	P Sample Percent Not Matched	E Sample Percent Erroneous Enumeration
Mail Out/Mail Back	8.5	4.7
Update/Leave	7.9	4.4
List/Enumerate	16.3	7.3
Rural Update/Enumerate	12.2	3.4
Urban Update/Leave	8.1	4.5
Urban Update/Enumerate	6.6	2.1
Adds to Address List	13.8	4.0
Total	8.6	4.6

The males in the age group 18 to 29 appear to have higher P sample percent not matched and E sample percent erroneous enumeration than other age and sex groups in Table 9e when both age and sex is given.

Table 9e: Percent Not Matched and Erroneously Enumerated After Follow-up by Age and Sex Before Weighting and Imputation

Age and Sex		P Sample Percent Not Matched	E Sample Percent Erroneous Enumeration
Under 18	Male and Female	9.2	3.0
18 to 29	Male	14.0	5.1
18 to 29	Female	11.4	4.6
30 to 49	Male	8.8	3.3
30 to 49	Female	7.0	2.8
50 and over	Male	6.1	3.4
50 and over	Female	5.5	3.4
Blank	Male	15.7	38.5
Blank	Female	12.6	39.1
18 to 29	Blank	18.8	11.8
30 to 49	Blank	11.5	8.2
50 and over	Blank	9.5	11.6
Blank	Blank	11.3	86.3
Total		8.6	4.6

The actual race domains used in the post stratification were not available on our data file when we made these tables. We simulated the race domain and created a category called multiple races when two or more races were marked. We also have a category for blank race because the imputation was not available. People in the white race domain appear to have lower P sample percent not matched and E sample percent erroneous enumeration than people in the other race domains in Table 9f.

Table 9f: Percent Not Matched and Erroneously Enumerated After Follow-up by Simulated Race Domain Before Weighting and Imputation

Simulated Race Domain	P Sample Percent Not Matched	E Sample Percent Erroneous Enumeration
Blank	16.6	6.6
American Indians on Reservations (including Hispanics and multi-race people living on reservations)	12.4	3.0
American Indians Not on Reservations (Hispanic and non-Hispanic)	12.2	5.0
Hispanic alone	12.4	7.3
Black alone	13.2	6.4
Pacific Islander (Hispanics, non- Hispanic and multi-race people living in Hawaii)	15.1	3.7
Asian alone	9.6	4.5
White alone	6.3	3.5
Other Races alone	11.2	6.8
Multiple Races	9.3	3.5
Total	8.6	4.6

People in the high mail return rate category appear to have lower P sample percent not matched and E sample percent erroneous enumeration than people in the low mail return rate category in Table 9g.

Table 9g: Percent Not Matched and Erroneously Enumerated After Follow-up by Mail Return Rate Before Weighting and Imputation

Mail Return Rate	P Sample Percent Not Matched	E Sample Percent Erroneous Enumeration
Greater than 25 percent	6.7	3.8
Less than 25 percent	14.4	7.6
List/Enumerate, Rural Update Enumerate, Urban Update/Enumerate	12.9	4.0
No Occupied Housing Units or Incomplete Address	28.7	0.0
Total	8.6	4.6

We did not have housing variables in the census to indicate multi-units and mobile homes. We created variables by classifying the clusters based on the housing unit variables from the A.C.E. listing. The next two tables contain percent not matched and erroneous enumeration for clusters with different amounts of multi-units and mobile homes. People in clusters with high rates of mobile homes appear to have higher P sample percent not matched and E sample percent erroneous enumeration than people in clusters with fewer mobile homes in Table 9h.

Table 9h: Percent Not Matched and Erroneously Enumerated
After Follow-up by Percent Mobile Home Before Weighting and
Imputation

Percent Mobile Home	P Sample Percent Not Matched	E Sample Percent Erroneous Enumeration
None	8.4	4.6
10 Percent or less	7.9	4.4
11 to 50 percent	9.1	4.3
Greater than 50 percent	12.1	6.5
Total	8.6	4.6

People in clusters high rates of multi-units appear to have higher P sample percent not matched and E sample percent erroneous enumeration than people in clusters with fewer multi-units in Table 9i.

Table 9i: Percent Not Matched and Erroneously Enumerated After Follow-up by Percent Multi-Unit Before Weighting and Imputation

Percent Multi-Unit	P Sample Percent Not Matched	E Sample Percent Erroneous Enumeration
None	7.3	3.9
10 Percent or less	7.0	3.8
11 to 50 percent	8.5	4.9
Greater than 50 percent	12.7	6.6
Total	8.6	4.6

People identified as nonmovers appear to have lower P sample percent not matched than people identified as outmovers in Table 9j.

Table 9j: Percent Not Matched After Follow-up by A.C.E. Mover Status Before Weighting and Imputation

Mover Status	P Sample Percent Not Matched
Nonmover	8.0
Outmover	23.6
Unresolved Mover Status	22.6
Total	8.6

P sample people interviewed by telephone appear to have lower P sample percent not matched than people interviewed by personal visit in Table 9k.

Table 9k: Percent Not Matched After Follow-up by
A.C.E. Interview Mode Before Weighting and
Imputation

Interview Mode	P Sample Percent Not Matched
Telephone	1.4
Personal Visit	12.5
Quality Assurance Replacement	12.9
Total	8.6

P sample people interviewed by proxy appear to have higher P sample percent not matched than people interviewed by household member in Table 9l.

Table 9l: Percent Not Matched After Follow-up by
A.C.E. Respondent Type Before Weighting and
Imputation

Respondent Type	P Sample Percent Not Matched
Household Member	8.1
Proxy Respondent	21.3
Total	8.6

What are the types of erroneous enumerations?

Erroneously enumerated - The categories were people with insufficient information for matching and follow-up, duplicates, fictitious, geocoding errors, and people who should have been enumerated at another residence on census day.

- The E sample people with insufficient information for matching and follow-up were data-defined, but did not contain full name and at least two characteristics.
- The E sample people enumerated more than once were coded as duplicates.

- The fictitious people are ones where we found notes on the census image identifying the person as not a real person such as a dog or other pet or they were identified as not existing in this cluster during the follow-up interview. Three respondents who never heard of the person were required in order to code a person as fictitious.
- Census people in housing units identified as geocoding errors during the housing unit follow-up were coded as erroneously enumerated because of geocoding error.
- The E sample person should have been counted at another residence on census day.

The next table contains the final weighted and imputed data after person follow-up in 1990 for the erroneous enumerations. See Hogan (1993).

Table 10: 1990 Erroneous Enumerations Final Weighted Numbers

E Sample Erroneous Enumeration Code	Percent of Erroneous Enumerations	Percent of E Sample
Insufficient Information	20.8	1.2
Duplicate	28.2	1.6
Fictitious	2.6	0.2
Geocoding Error	6.0	0.3
Other Residence	38.0	2.2
Unresolved	4.5	0.3
Total	100.0	5.8

The percentages of each type of erroneous enumeration in these tables were based on the E sample people with a resolved enumeration status. Note that the percentage of each type of erroneous enumeration is one hundred times the rate of each type of erroneous enumeration. The percent duplicate includes the duplication between E sample and census people not in the E sample after subsampling large clusters. Duplicates between the E sample and non E sample people in the cluster are not whole erroneous enumerations. A probability of erroneous enumeration caused by duplication is calculated for the E sample person duplicated to a non E sample person within the block cluster.

$$\text{Rate of Type of Erroneous Enumeration} = \frac{\text{Type of Erroneous Enumeration}}{\text{Correct Enumeration} + \text{Erroneous Enumeration}}$$

Table 11a: 2000 Unweighted Types of
Erroneous Enumerations

E Sample Erroneous Enumeration Code	Percent of E Sample
Insufficient Information	2.0
Duplicate	0.9
Fictitious	0.3
Geocoding Error	0.3
Other Residence	1.1
Total	4.6

The remaining tables contain the type of erroneous enumeration as a percent of the total E sample resolved cases by different variables. These are unweighted numbers with the people not in sample for the targeted extended search removed.

The actual race domains used in the post stratification were not available on our data files when we made these tables. We simulated the race domain and created a category called multiple races when two or more races were marked. We also have a category for blank race because the imputation was not available. The Hispanic and Black race domains had high percentages of insufficient information for matching and follow-up and duplicates in Table 11b. The American Indians on Reservations had higher rates of other residence.

Table 11b: Percent of E Sample for Type of Erroneous Enumeration After Follow-up Before Weighting and Imputation by Simulated Race Domain

Simulated Race Domain	Insufficient Information	Duplicate	Fictitious	Geocoding Error	Other Residence
Blank	3.6	1.2	0.3	0.2	1.3
American Indians on Reservations (including Hispanics and multi-race people living on reservations)	0.6	0.7	0.1	0.0	1.6
American Indians Not on Reservations (Hispanic and non-Hispanic)	2.3	1.1	0.2	0.2	1.2
Hispanic alone	4.0	1.4	0.4	0.4	1.1
Black alone	2.8	1.4	0.8	0.3	1.2
Pacific Islander (Hispanics, non-Hispanic and multi-race people living in Hawaii)	1.3	1.1	0.5	0.0	0.8
Asian alone	1.5	1.1	0.4	0.3	1.3
White alone	1.3	0.7	0.2	0.3	1.1
Other Races alone	2.7	1.6	0.7	0.6	1.2
Multiple Races	1.3	0.6	0.2	0.2	1.1
Total	2.0	0.9	0.3	0.3	1.1

Table 11c: Percent of E Sample for Type of Erroneous Enumeration After Follow-up Before Weighting and Imputation by Type of Enumeration Area

Type of Enumeration Area	Insufficient Information	Duplicate	Fictitious	Geocoding Error	Other Residence
Mail Out/Mail Back	2.1	0.9	0.3	0.3	1.0
Update/Leave	1.5	1.2	0.1	0.0	1.6
List/Enumerate	2.0	0.4	0.0	4.0	0.9
Rural Update/Enumerate	1.0	0.6	0.1	0.1	1.5
Urban Update/Leave	3.0	0.3	0.2	0.1	0.8
Urban Update/Enumerate	0.6	0.0	0.0	0.0	1.5
Adds to Address List	1.5	1.0	0.0	0.0	1.5
Total	2.0	0.9	0.3	0.3	1.1

The two categories of age 18 to 29 had high percentages of other residence in Table 11d. The percentages of insufficient information for matching and follow-up, duplicate, fictitious, and geocoding error seemed fairly consistent for the age and sex groups with complete data.

Table 11d: Percent of E Sample for Type of Erroneous Enumeration After Follow-up Before Weighting and Imputation by Age and Sex

Age and Sex		Insufficient Information	Duplicate	Fictitious	Geocoding Error	Other Residence
Under 18	Male and Female	0.9	0.7	0.3	0.3	0.8
18 to 29	Male	0.9	1.1	0.5	0.2	2.2
18 to 29	Female	0.9	1.0	0.4	0.3	1.9
30 to 49	Male	0.9	0.8	0.4	0.3	0.9
30 to 49	Female	0.8	0.7	0.3	0.3	0.6
50 and over	Male	0.8	0.9	0.2	0.3	1.2
50 and over	Female	1.0	0.9	0.1	0.3	1.1
Blank	Male	30.9	4.0	0.7	0.3	2.6
Blank	Female	31.4	4.2	0.8	0.3	2.5
18 to 29	Blank	51.1	1.6	1.5	0.3	3.3
30 to 49	Blank	4.8	1.9	0.6	0.1	0.8
50 and over	Blank	7.0	2.2	0.2	0.3	2.0
Blank	Blank	83.0	2.4	0.2	0.0	0.8
Total		2.0	0.9	0.3	0.3	1.1

The renters had high percentages of insufficient information for matching and follow-up, duplication, and fictitious than the owners in Table 11e.

Table 11e: Percent of E Sample for Type of Erroneous Enumeration After Follow-up Before Weighting and Imputation by Tenure

Tenure	Insufficient Information	Duplicate	Fictitious	Geocoding Error	Other Residence
Owner	1.0	0.6	0.1	0.3	1.0
Renter	2.4	1.4	0.6	0.3	1.2
Blank	13.9	2.3	0.6	0.3	2.0
Total	2.0	0.9	0.3	0.3	1.1

The people in the Northeast appear to have higher percentages of duplication in Table 11f.

Table 11f: Percent of E Sample for Type of Erroneous Enumeration After Follow-up Before Weighting and Imputation by Region

Region	Insufficient Information	Duplicate	Fictitious	Geocoding Error	Other Residence
Northeast	2.0	1.5	0.4	0.2	1.1
Midwest	1.6	0.7	0.2	0.2	0.9
South	2.0	0.9	0.3	0.4	1.3
West	2.1	0.7	0.2	0.4	1.1
Total	2.0	0.9	0.3	0.3	1.1

The people in the New York Regional Office appear to have higher percentages of insufficient information for matching and follow-up and duplication in Table 11g.

Table 11g: Percent of E Sample for Type of Erroneous Enumeration After Follow-up Before Weighting and Imputation by Regional Office

Regional Office	Insufficient Information	Duplicate	Fictitious	Geocoding Error	Other Residence
Boston	1.5	1.3	0.2	0.3	1.3
New York	2.7	2.6	0.7	0.1	1.0
Philadelphia	2.2	0.8	0.3	0.2	1.0
Detroit	1.6	0.5	0.2	0.2	0.9
Chicago	1.8	0.9	0.4	0.3	0.8
Kansas City	1.5	0.7	0.2	0.0	1.1
Seattle	2.5	0.9	0.2	0.4	1.1
Charlotte	1.8	0.8	0.3	0.1	1.2
Atlanta	2.3	0.8	0.2	0.7	1.6
Dallas	1.9	1.0	0.5	0.4	1.2
Denver	1.7	0.6	0.2	0.4	1.4
Los Angeles	2.0	0.7	0.2	0.3	0.8
Total	2.0	0.9	0.3	0.3	1.1

The people in the clusters with more mobile homes appear to have higher percentages of insufficient information for matching and follow-up and duplication in Table 11h.

Table 11h: Percent of E Sample for Type of Erroneous Enumeration After Follow-up Before Weighting and Imputation by Percent Mobile Homes in the Cluster

Percent Mobile Home	Insufficient Information	Duplicate	Fictitious	Geocoding Error	Other Residence
None	2.1	0.9	0.3	0.3	1.0
10 Percent or less	1.7	0.9	0.2	0.3	1.3
11 to 50 percent	1.5	1.1	0.2	0.2	1.4
Greater than 50 percent	2.1	1.5	0.2	0.6	2.0
Total	2.0	0.9	0.3	0.3	1.1

The people in the clusters with more multi-units appear to have higher percentages of insufficient information for matching and follow-up and duplication in Table 11i.

Table 11i: Percent of E Sample for Type of Erroneous Enumeration After Follow-up Before Weighting and Imputation by Percent Multi-Unit in the Cluster

Percent Multi-Unit	Insufficient Information	Duplicate	Fictitious	Geocoding Error	Other Residence
None	1.6	0.6	0.2	0.4	1.1
10 Percent or less	1.6	0.7	0.2	0.2	1.1
11 to 50 percent	2.1	1.2	0.3	0.2	1.1
Greater than 50 percent	3.0	1.7	0.6	0.2	1.1
Total	2.0	0.9	0.3	0.3	1.1

The people in the white and other, owner, and with low mail return categories in both mail out/mail back and other types of enumeration areas in the Northeast appear to have higher percentages of duplication in Table 11j. Also, people in the Hispanic, renter, not in mail out/mail back type of enumeration area, and low mail return rate appear to have high rates of insufficient information for matching and follow-up and of duplication.

Table 11j: Percent of E Sample for Type of Erroneous Enumeration After Follow-up Before Weighting and Imputation for Simulated Post-Strata Variables

Simulated Race Domain	Tenure	TEA	Mail Return Rate	Region	Insufficient Information	Duplicate	Fictitious	Geocoding Error	Other Residence
White and Other	Owner	MO/MB	High	Northeast	0.6	0.4	0.1	0.2	0.1
				Midwest	0.6	0.2	0.1	0.3	0.7
				South	0.8	0.3	0.1	0.2	1.1
				West	1.0	0.3	0.0	0.3	0.9
	Low			Northeast	1.0	2.2	0.9	0.9	1.1
				Midwest	1.1	1.3	0.1	0.1	1.0
				South	1.1	1.2	0.4	1.4	1.2
				West	1.4	1.5	0.2	0.8	1.4
	Other TEA	High		Northeast	0.5	0.5	0.0	0.2	1.2
				Midwest	0.4	0.4	0.0	0.0	1.1
				South	0.5	0.7	0.1	0.0	1.4
				West	0.9	0.3	0.1	1.0	1.7
	Low			Northeast	0.7	2.3	0.0	0.0	1.7
				Midwest	0.3	1.3	0.1	0.0	1.5
				South	0.7	1.6	0.1	0.0	1.3
				West	1.7	1.6	0.2	0.1	2.9

Race Domain	Tenure	TEA	Mail Return Rate	Region	Insufficient Information	Duplicate	Fictitious	Geocoding Error	Other Residence
White and Other	Renter	MO/MB	High		1.9	0.9	0.3	0.2	1.1
			Low		2.6	2.3	0.6	0.5	1.2
			High		1.2	0.8	0.1	0.5	1.8
	Owner	MO/MB	Low		1.5	2.3	0.6	0.0	1.7
			High		1.2	0.4	0.4	0.6	1.0
			Low		2.1	3.1	0.5	0.1	1.0
Black	Renter	MO/MB	High		0.8	1.0	0.1	0.0	2.0
			Low		0.3	2.4	0.1	0.0	0.9
			High		2.8	0.7	1.1	0.2	1.4
	Owner	MO/MB	Low		3.2	1.8	1.2	0.2	1.0
			High		0.8	1.2	1.2	0.1	1.0
			Low		1.5	1.4	0.4	0.0	1.6
Hispanic	Renter	MO/MB	High		2.0	0.5	0.1	0.2	0.8
			Low		1.9	2.2	0.8	0.2	1.1
			High		2.3	0.9	0.1	0.2	1.5
	Owner	MO/MB	Low		2.1	2.8	0.1	0.0	1.6
			High						
			Low						

Race Domain	Tenure	TEA	Mail Return Rate	Region	Insufficient Information	Duplicate	Fictitious	Geocoding Error	Other Residence
	Renter	MO/MB	High		3.0	1.3	0.4	0.3	0.9
			Low		3.6	2.3	1.0	1.1	1.2
		Other TEA	High		4.2	1.2	0.2	0.2	2.1
			Low		5.0	5.5	0.7	0.1	1.1
Pacific Islander	Owner				0.8	1.0	0.4	0.0	0.6
	Renter				1.0	1.1	0.5	0.1	0.9
Asian	Owner				0.9	0.9	0.2	0.3	1.1
	Renter				1.8	1.2	0.7	0.2	1.5
American Indians on Reservations	Owner				0.4	0.7	0.1	0.1	1.6
	Renter				0.5	0.3	0.1	0.0	1.7
American Indians Not on Reservations	Owner				2.1	0.8	0.1	0.2	1.1
	Renter				1.8	1.2	0.4	0.2	1.1
Multiple Races	Owner				1.0	0.6	0.0	0.2	1.1
	Renter				1.5	0.5	0.5	0.3	1.0
Domain or Tenure Blank					12.3	2.1	0.6	0.3	1.9

What are the unresolved rates?

The unresolved codes for the P sample were either unresolved match status or unresolved residence status. Cases with unresolved residence status are: matches with unsuccessful follow-up interview and nonmatches with unsuccessful follow-up interviews. Cases with both unresolved match and unresolved residence status are: possible matches with unsuccessful follow-up interview and the P sample people with insufficient information for matching and follow-up. Tables 12a through 12l contain the percent of the total P sample with unresolved match or residence status. All people with unresolved match status also have unresolved residence status, so the percentage with unresolved match status or unresolved residence status was the percentage of unresolved residence status. The percentage with unresolved match or residence status was one hundred times the rate of unresolved match or residence status.

$$\begin{array}{l} \text{Rate of Unresolved Match} \\ \text{or Residence Status} \end{array} = \frac{\text{People with Unresolved Residence Status}}{\text{Total P Sample People}}$$

Cases with unresolved enumeration status for the E sample are the ones with unsuccessful follow-up interview. Tables 12a through 12l also contain the percent of the total E sample with unresolved enumeration status. The percentage with unresolved enumeration status is one hundred times the rate of unresolved enumeration status.

$$\begin{array}{l} \text{Rate of Unresolved} \\ \text{Enumeration Status} \end{array} = \frac{\text{People with Unresolved Enumeration Status}}{\text{Total E sample People}}$$

Table 12a: Percent Unresolved After Follow-up by Census Region Before Weighting and Imputation

Census Region	P Sample Percent Unresolved	E Sample Percent Unresolved
Northeast	1.9	2.2
Midwest	1.8	2.2
South	2.7	3.6
West	2.7	3.5
Total	2.4	3.0

Table 12b: Percent Unresolved After Follow-up by Census Regional Office Before Weighting and Imputation

Census Regional Office	P Sample Percent Unresolved	E Sample Percent Unresolved
Boston	0.8	1.3
New York	3.2	3.1
Philadelphia	2.9	3.3
Detroit	1.6	1.9
Chicago	2.1	2.3
Kansas City	1.8	2.5
Seattle	2.7	3.0
Charlotte	2.8	3.8
Atlanta	2.9	3.0
Dallas	2.2	3.7
Denver	2.3	3.0
Los Angeles	2.9	4.2
Total	2.4	3.0

Table 12c: Percent Unresolved After Follow-up by Tenure Before Weighting and Imputation

Tenure	P Sample Percent Unresolved	E Sample Percent Unresolved
Owner	1.3	1.3
Renter	4.0	6.1
Blank	8.4	4.0
Total	2.4	3.0

Table 12d: Percent Unresolved After Follow-up by Type of Enumeration Area Before Weighting and Imputation

Type of Enumeration Area	P Sample Percent Unresolved	E Sample Percent Unresolved
Mail Out/Mail Back	2.5	3.2
Update/Leave	1.7	1.8
List/Enumerate	3.7	12.4
Rural Update/Enumerate	1.9	1.6
Urban Update/Leave	1.9	1.4
Urban Update/Enumerate	1.2	1.4
Adds to Address List	1.5	0.9
Total	2.4	3.0

Table 12e: Percent Unresolved After Follow-up by Age and Sex Before Weighting and Imputation

Age and Sex		P Sample Percent Unresolved	E Sample Percent Unresolved
Under 18	Male and Female	2.1	2.7
18 to 29	Male	3.1	6.2
18 to 29	Female	2.7	5.5
30 to 49	Male	1.9	3.1
30 to 49	Female	1.5	2.4
50 and over	Male	1.1	1.8
50 and over	Female	1.0	1.5
Blank	Male	16.9	6.2
Blank	Female	14.8	5.2
18 to 29	Blank	8.1	6.0
30 to 49	Blank	4.5	3.8
50 and over	Blank	3.8	2.1
Blank	Blank	69.9	0.6
Total		2.4	3.0

Table 12f: Percent Unresolved After Follow-up by Simulated Race Domain Before Weighting and Imputation

Simulated Race Domain	P Sample Percent Unresolved	E Sample Percent Unresolved
Blank	16.6	3.2
American Indians on Reservations (including Hispanics and multi-race people living on reservations)	1.7	1.6
American Indians Not on Reservations (Hispanic and non- Hispanic)	2.8	4.2
Hispanic alone	4.7	3.9
Black alone	3.1	4.2
Pacific Islander (Hispanics, non- Hispanic and multi-race people living in Hawaii)	1.5	3.5
Asian alone	2.4	3.6
White alone	1.6	2.5
Other Races	2.3	4.6
Multiple Races	1.5	3.3
Total	2.4	3.0

Table 12g: Percent Unresolved After Follow-up by Mail Return Rate Before Weighting and Imputation

Mail Return Rate	P Sample Percent Unresolved	E Sample Percent Unresolved
Greater than 25 percent	2.0	2.3
Less than 25 percent	3.5	5.3
List/Enumerate, Rural Update Enumerate, Urban Update/Enumerate	2.3	3.8
No Occupied Housing Units or Incomplete Address	5.7	0.0
Total	2.4	3.0

We do not have housing variables in the census to indicate multi-units and mobile homes. We created variables by classifying the clusters based on the housing unit variables from the A.C.E. listing. The next two tables contain percent not matched and erroneous enumeration for clusters different amounts of multi-units and mobile homes.

Table 12h: Percent Unresolved After Follow-up by Percent Mobile Home Before Weighting and Imputation

Percent Mobile Home	P Sample Percent Unresolved	E Sample Percent Unresolved
None	2.5	3.2
10 Percent or less	1.7	2.3
11 to 50 percent	1.8	1.9
Greater than 50 percent	2.7	3.4
Total	2.4	3.0

Table 12i: Percent Unresolved After Follow-up by
Percent Multi-Unit Before Weighting and Imputation

Percent Multi-Unit	P Sample Percent Unresolved	E Sample Percent Unresolved
None	1.9	2.2
10 Percent or less	1.8	2.0
11 to 50 percent	2.2	2.9
Greater than 50 percent	3.9	5.5
Total	2.4	3.0

Table 12j: Percent Unresolved After Follow-up by
A.C.E. Mover Status Before Weighting and Imputation

Mover Status	P Sample Percent Unresolved
Nonmover	1.4
Outmover	15.2
Unresolved Mover Status	36.2
Total	2.4

Table 12k: Percent Unresolved After Follow-up by A.C.E. Interview Mode Before Weighting and Imputation

Interview Mode	P Sample Percent Unresolved
Telephone	0.8
Personal Visit	3.2
Quality Assurance Replacement	4.4
Total	2.4

Table 12l: Percent Unresolved After Follow-up by A.C.E. Respondent Type Before Weighting and Imputation

Respondent Type	P Sample Percent Unresolved
Household Member	1.6
Proxy Respondent	19.2
Total	2.4

What are the final census day interview outcome codes?

The final census day outcome codes are in Tables 13. Changes as a result of the follow-up interview are

- Whole households of P sample people who said they lived elsewhere on census day were converted to a noninterviews.
- Whole households who lived in group quarters on census day or should have been enumerated at another residence were converted to vacant.

Table 13a: Final Census Day Interviewing Outcome for A.C.E. Housing Units

Outcome Code	Unweighted Housing Units	Percent
Interview		
Complete interview with a household member	233,327	77.5
Complete interview with a proxy respondent	18,335	6.1
Sufficient partial interview	2,513	0.8
Noninterview		
No census day residents	2,709	0.9
Field noninterview	2,667	0.9
All people have insufficient information for matching and follow-up	2,418	0.8
Vacant		
No census day residents	4,561	1.5
Vacant on census day	23,911	8.0
Not a housing unit		
Not a housing unit on census day	10,472	3.7
Total	300,913	100.0

Addresses that were not housing units on census day were removed from the P sample. The next table contains numbers of housing units identified as interviews, noninterviews, and vacant and percents of total housing units and numbers and percents of occupied housing units. The noninterview rate for occupied housing units for census day is 3.0 percent

Table 13b: Final Census Day Interviewing Outcome for P Sample Housing Units

Outcome Code	Total P Sample Housing Units		Occupied P Sample Housing Units	
	Unweighted Housing Units	Percent	Unweighted Housing Units	Percent
Interview	254,175	87.5	254,175	97.0
Noninterview	7,794	2.7	7,794	3.0
Vacant	28,472	9.8		
Total	290,441	100.0	261,969	100.0

In 1990, 1.6 percent of the occupied PES housing units were noninterviews. Procedure B, which searches for the current residents at their census day address, was used in 1990. The percent noninterview is for the current residents which include the nonmovers and in-movers. See Hogan (1993).

Table 14: 1990 P Sample Housing Units

	Total Housing Units	Occupied Housing Units
Interviews	85.2	98.4
Household Member	81.2	93.7
Non Household Member	4.1	4.7
Noninterviews	1.4	1.6
Occupied Housing Units	86.6	100.0
Vacant	13.4	
Total Housing Units	100.0	

What are census day noninterview rates for occupied housing units?

The census day noninterview rates in the next set of tables are for occupied housing units. The interviewed housing units and the noninterviewed housing units are added together yielding the total number of occupied housing units.

$$\text{Noninterview Rates for Occupied Housing Units} = \frac{\text{Noninterviews}}{\text{Interviews} + \text{Noninterviews}}$$

These data are unweighted sample data.

The census day noninterview rates are recalculated to reflect changes due to coding in after follow-up matching.

Table 15a: P Sample Noninterview Rates for Census Day in Occupied Housing Units by Interview Mode (Unweighted Data)

Interview Mode	Percent Noninterview
Telephone	1.1
Personal	3.7
Quality Assurance	37.4

Table 15b: P Sample Noninterview Rates for Census Day in Occupied Housing Units by Census Regional Office (Unweighted Data)

Census Regional Office	Percent Noninterview
Boston	1.2
New York	4.2
Philadelphia	3.6
Detroit	2.3
Chicago	2.7
Kansas City	2.3
Seattle	2.6
Charlotte	4.3
Atlanta	3.6
Dallas	3.4
Denver	2.6
Los Angeles	2.8
Total	3.0

Table 15c: P Sample Noninterview Rates for Census Day
in Occupied Housing Units by Census Region
(Unweighted Data)

Census Region	Percent Noninterview
Northeast	2.6
Midwest	2.4
South	3.7
West	2.8

Table 15d: P Sample Noninterview Rates for Census Day
in Occupied Housing Units by Type of Enumeration Area
(Unweighted Data)

Type of Enumeration Area	Percent Noninterview
Mail Out / Mail Back	3.0
Update/Leave	2.9
List/Enumerate	2.3
Rural Update/Enumerate	4.0
Urban Update/Leave	2.9
Urban Update/Enumerate	3.5
Additions to Address Listing	3.1

Table 15e: P Sample Noninterview Rates for Census Day
in Occupied Housing Units by Whether the Interview was
done in the Nonresponse Conversion Operation (NRCO)
(Unweighted Data)

	Percent Noninterview
Interview not done in NRCO	2.7
Interview obtained in NRCO	10.4

Table 15f: P Sample Noninterview Rates for Census Day in Occupied Housing Units by Type of Structure from A.C.E. Listing (Unweighted Data)

Type of Structure	Percent Noninterview
Single Unit	2.1
Multi-Unit	4.9
Mobile Home not in Park	3.9
Mobile Home in Park	4.4
Single Unit in a Special Place	2.5
Multi-unit in a Special Place	5.5
Other	9.8

Table 15g: P Sample Noninterview Rates for Census Day in Occupied Housing Units by Status of A.C.E. Listing(Unweighted Data)

Status of A.C.E. Listing	Percent Noninterview
Housing Unit	2.9
Under Construction	4.5
Future Construction	9.0
Unfit for Habitation	24.3
Boarded Up	16.2
Storage	17.8
Vacant Trailer Site	8.5
Other	5.1

Table 15h: P Sample Noninterview Rates for Census Day
in Occupied Housing Units
by Type of Interview (Unweighted Data)

Type of Interview	Percent Noninterview
Interview with a Household Member	1.8
Proxy Interview	17.4

How do the preliminary and final census day interview outcome codes compare?

Table 16 compares the preliminary and final census day interview outcome codes. Changes in preliminary and final census day outcome codes occur for the people identified as interviewed for a preliminary outcome code. The follow-up interview for P sample persons can identify them as not a resident of the housing unit because they did not live at the sample address or lived at the sample address, but should have been counted at another residence such as group quarters or another home. The housing unit can also be identified as not being a housing unit on census day.

Table 16: Comparison of the Preliminary and Final Census Day Outcome Codes

Preliminary Census Day Outcome Codes	Final Census Day Outcome Codes								
	Interview with Household Member	Interview with Proxy	Partial	No Census Day Residents - Noninterview	Field Non- interview	Whole Household Insufficient Information	No Census Day Residents - Vacant	Vacant	Not a Housing Unit
Interview with Household Member	233,327	0	0	2,033	0	0	125	0	147
Interview with Proxy	0	18,335	0	676	0	0	252	0	117
Partial Interview	0	0	2,513	0	0	97	0	0	2
Field Noninterview	0	0	0	0	2,667	0	0	0	0
Whole Household Insufficient Information	0	0	0	0	0	2,321	0	0	0
No Census Day Residents	0	0	0	0	0	0	4,184	0	0
Vacant	0	0	0	0	0	0	0	23,911	0
Not a Housing Unit	0	0	0	0	0	0	0	0	10,206

What are the final interview day interview outcome codes?

The final interview day outcome codes are in Table 17. The interview outcome as of interview day is for the nonmovers and the in-movers. Changes as a result of the follow-up interview are in whole households of nonmovers who said they lived elsewhere, in group quarters, or have another residence where they should have been counted on census day are converted to noninterviews.

Table 17: Final Interview Day Estimation Outcome Codes for P Sample Housing Units (Unweighted Data)

Outcome Code	Housing Units	Percent
Interview		
Complete interview with a household member	249,854	83.0
Complete interview with a proxy respondent	12,317	4.1
Partial interview	1,932	0.6
Noninterview		
No census day residents - household converted to noninterview	483	0.2
Field noninterview	373	0.1
All people have insufficient information for matching and follow-up	2,196	0.7
Vacant		
Vacant on census day	29,662	9.9
Not a housing unit		
Not a housing unit on census day	4,096	1.4
Total	300,913	100.0

Table 18: Final Interview Day Interviewing Outcome for P Sample Housing Units

Outcome Code	Total P Sample Housing Units		Occupied P Sample Housing Units	
	Unweighted Housing Units	Percent	Unweighted Housing Units	Percent
Interview	264,103	89.0	264,103	98.9
Noninterview	3,052	1.0	3,052	1.1
Vacant	29,662	10.0		
Total	296,817	100.0	267,155	100.0

The interview day noninterview rates are recalculated to reflect changes due to coding in after follow-up matching. The final noninterview rates for interview day are in the next set of tables.

Table 19a: P Sample Noninterview Rates for Interview Day in Occupied Housing Units by Interview Mode (Unweighted Data)

Interview Mode	Percent Noninterview
Telephone	0.7
Personal	1.0
Quality Assurance	15.4

Table 19b: P Sample Noninterview Rates for Interview Day
in Occupied Housing Units by Census Regional Office
(Unweighted Data)

Census Regional Office	Percent Noninterview
Boston	0.1
New York	1.5
Philadelphia	1.2
Detroit	0.6
Chicago	1.1
Kansas City	0.6
Seattle	1.0
Charlotte	1.8
Atlanta	1.0
Dallas	1.0
Denver	0.6
Los Angeles	1.1
Total	1.1

Table 19c: P Sample Noninterview Rates for Interview Day
in Occupied Housing Units
by Census Region (Unweighted Data)

Census Region	Percent Noninterview
Northeast	0.8
Midwest	0.8
South	1.3
West	0.9

Table 19d: P Sample Noninterview Rates for Interview Day
in Occupied Housing Units by Type of Enumeration Area
(Unweighted Data)

Type of Enumeration Area	Percent Noninterview
Mail Out / Mail Back	1.1
Update/Leave	0.7
List/Enumerate	0.3
Rural Update/Enumerate	0.4
Urban Update/Leave	0.5
Urban Update/Enumerate	0.9
Additions to Address Listing	0.0

Table 19e: P Sample Noninterview Rates for Interview Day
in Occupied Housing Units by Whether the Interview was
done in the Nonresponse Conversion Operation (NRCO)
(Unweighted Data)

	Percent Noninterview
Interview not done in NRCO	0.8
Interview obtained in NRCO	5.7

Table 19f: P Sample Noninterview Rates for Interview Day in Occupied Housing Units by Type of Structure from A.C.E. Listing (Unweighted Data)

Type of Structure	Percent Noninterview
Single Unit	0.8
Multi-Unit	1.5
Mobile Home not in Park	0.6
Mobile Home in Park	1.0
Single Unit in a Special Place	0.7
Multi-unit in a Special Place	1.8
Other	1.9

Table 19g: P Sample Noninterview Rates for Interview Day in Occupied Housing Units by Status of A.C.E. Listing (Unweighted Data)

Status of A.C.E. Listing	Percent Noninterview
Housing Unit	1.0
Under Construction	0.7
Future Construction	1.0
Unfit for Habitation	2.1
Boarded Up	0.5
Storage	1.3
Vacant Trailer Site	0.7
Other	1.4

Table 19h: P Sample Noninterview Rates for Interview Day in Occupied Housing Units by Type of Interview (Unweighted Data)

Type of Interview	Percent Noninterview
Interview with a Household Member	0.5
Proxy Interview	8.6

Quality Assurance of the Clerical Person Matching Operation

How was quality assurance operationalized?

The quality assurance on clerical matching was based on a three-tiered dependent review of person records.

How was the clerical matching performed?

- We computer matched the P sample to the census using the Census Statistical Research Division Record Linkage System.
- Clerical personnel at the National Processing Center (NPC) reviewed records that were not matched by the computer matcher.
- There were 225 clerks, 46 technicians, and 16⁵ analysts. Each successive level performed quality assurance (QA) on the previous level.
- A higher level user dependently reviewed a sample of each user's work, a process that should identify random matching errors. Each of the matching levels improved on the previous level. The clerks matched what the computer matcher could not. The technicians worked on any cases the clerks could not resolve and performed the quality assurance on the clerks. Then the analysts finished any cases the technicians could not resolve and performed the quality assurance on the technicians. Clusters with match results that required a second opinion were sent to the higher stage.
- Users worked records in clusters during before followup (BFU). In after followup (AFU), users worked records in batches of person followup (PFU) forms that were irrespective of clusters. (See Assumptions section for more information).

⁵There are 10 production analysts and 6 managers and headquarters personnel that worked as analysts.

What was the quality assurance plan?

The QA plan was designed to control the quality of the clerical matching by targeting both records that required a higher level of review and individual matchers who required more consistent review, and to keep a uniform workflow throughout matching.

- The BFU QA plan was aimed primarily at finding errors that could have prevented people who should have been followed up from being assigned to the followup operation.
- The AFU QA plan was aimed at finding errors that could have caused people to have been incorrectly classified as either correctly or erroneously enumerated, as well as errors that could have caused people to incorrectly be classified as removes from the P sample.

What were the quality assurance operations?

Three levels of users worked the clusters and batches dependently, to ensure that coding errors would be kept to a minimum.

- The Person Matching Review and Coding System (PERMaRCS) monitored the work of all matchers to target matchers who required a more consistent review.
- All users (clerks and technicians) began with either 100 percent of their work being reviewed or began with a sample of their work being reviewed. This determination was made based on results from a non-production, predetermined set of clusters that were worked by each user before production began.
- PERMaRCS approved users to be reviewed on a sample basis who began in 100 percent review after completing 200 records in BFU, then 100 records in AFU, with a change rate below the prespecified four percent.
- PERMaRCS monitored each clerk's and technician's matching results through the entire matching process by counting significant changes⁶ of codes.
- Each time a matcher worked 50 records that were reviewed by a higher level, the system reassessed that matcher's sampling status. Based on a change rate cutoff of four percent, matchers with a higher change rate were placed back into 100 percent review while matchers with a lower change rate were put into sampling mode. In this way, the sampling QA decision was periodically and automatically revisited for each matcher.

⁶Most changes were considered significant, however a review code could be entered by the clerk or technician to flag a record for review by the higher level technician or analyst. In this case, any code change was not considered significant. In other cases the distinction between the codes and the resulting damage if incorrectly coding a record with these codes was considered insignificant and the change was not counted.

- To target records that required a higher level of review, the system allowed users to flag difficult records for the next level of review.
- The system also checked for certain predefined situations and, if present, the cluster or batch was automatically sent to a higher level matcher for review. For example, if a technician changed more than half of a clerk's codes, an analyst reviewed all of the coding for that clerk and that technician.
- The matching software was designed to continuously check the matcher's work to minimize many kinds of mistakes, such as assigning an invalid match code or leaving some records uncoded.

What are the assumptions of the QA Plan?

The QA Plan had several assumptions:

- The change rate overestimated the true error rate.
- Individual code changes did not always indicate errors. Additional experience and training may have led a matcher at a higher level to code a record differently. The QA plan assumed a negative correlation between a matcher whose coding was frequently changed and quality of coding, but not a one-to-one correspondence between code changes and errors.
- Certain entire clusters and records that users reviewed at a higher level were considered to be out-of-sample for QA purposes.
- Clusters or batches of the work were routed to the technician or analyst either for a complete review (for matchers in 100 percent review or in samples selected for review), or because the cluster or batch contained a situation flagged by the system for higher review (e.g., certain Targeted Extended Search clusters where a clerk did not enter any Targeted Extended Search codes).
- Records in this latter category were usually indistinguishable from records that did not go to a higher review. In producing the statistics in this paper, we compensated by calculating our overall change rates using three different models, as shown below.
- The QA plan measured the quality of the clerks' and technicians' work, not the analysts.
- There was no QA of the analysts' work. Due to their extensive training and specific knowledge of the task, we assumed analysts have no errors.
- Each stage was considered individually, and the last two stages of AFU were disregarded for QA purposes.
- The rules for coding records in BFU were different than the rules for AFU and

consequently the types of changes were different. Therefore the AFU results could not be used to measure changes in BFU.

- Within the AFU stage, the rules and types of changes in the clerk, technician, and analyst stages were different from the coding rules and types of changes in cluster review and outlier stages. The first three stages reviewed the records individually (in batches) while the last two stages reconstructed the cluster for a cluster-level review. Therefore a change made to a record in the last two stages of AFU did not indicate an error in the first three stages of AFU. These changes were the result of additional information available during cluster review.
- Because the last two stages did not have QA, we disregarded any code change for the outgoing quality calculation. The last two stages of matching only improved the quality of matching by examining the cluster as a whole. The first cluster stage, cluster review, consisted of very targeted review that technicians performed. The second cluster stage, outlier review, was performed by analysts or technicians, in a very limited capacity.

What are the limitations on matching QA?

- The QA results presented in this document report only on the quality of the clerical matching stages in BFU and the first three stages of AFU.
- We computer matched 69.6 percent of the P sample and 64.4 percent of the E sample in the Computer Matching Phase using the Census Statistical Research Division Record Linkage System. The computer matcher assigned cutoffs very conservatively. Numerous studies over the years have shown that this operation was virtually error free (e.g., there were insignificant numbers of false matches).
- We included person records from Puerto Rico in these statistics. We sampled clusters and batches without regard to their geographic location.

What are the QA Results of the clerical matching operation?

The QA results of the clerical matching provided individual change rates in clusters/batches sampled for QA for clerks and technicians. These calculations could then be used to estimate the number of defects in records worked by clerks and technicians that did not get a higher review. For each stage of matching, the records were partitioned into the highest stage of review (BFU clerk, BFU technician, BFU analyst, AFU clerk, AFU technician, AFU analyst), and a resulting outgoing quality by stage was determined for the records completed in this stage (and not reviewed in any higher stage). Please see Appendix 2 for accompanying formulae.

What are the individual user change rates for technicians in sampled clusters/batches?

- We derived the BFU Technician individual change rate for each technician by looking at the records worked by that technician in clusters/batches sampled for QA. For each technician we divided the number of the user's records changed by

the analysts by the number of records checked by the analysts (Equation 1).

What are the individual change rates for clerks in sampled clusters/batches?

- We calculated the BFU Clerk individual change rate by determining the number of records worked by a given clerk that were changed by technicians or analysts in clusters/batches sampled for QA and then dividing this number by the total number of records worked by the clerk that were checked in clusters/batches sampled for QA (Equation 3).
- We assigned an adjusted overall technician change rate to records reviewed only by technicians to compensate for changes the technicians may have missed. To avoid overestimating the changes to the clerks' work, we used the adjusted rate to exclude those cases where the technician's code was changed, but the clerk properly coded the record (Equation 2).

How do we estimate the estimated overall change rates for BFU?

We used three different models to estimate the overall change rates in BFU for both clerks and technicians. For any given user, we classified records four ways:

- Randomly sampled for review (X)
- Not sampled for higher review, but part of a cluster that a higher level user worked, and the higher level user coded (Y')
- Not sampled for higher review, but part of a cluster that a higher level user worked, and a higher level user did not code (Y'')
- Not sampled for a higher review and not reviewed by a higher level (Z)

From the records in X, we had individual change rates, generalized here as p_x , for a given user. The sum of Y' and Y'' was Y, a cluster or batch that the system did not sample for higher review, but that a higher level matcher worked. Using the proportion p_x , we estimated the overall change rate (Equations 4 and 6) and outgoing quality (Equations 5 and 7) for the remaining records. For the clerk level, records were considered part of Y if a technician reviewed the cluster or batch, but the workunit was not sampled for QA. For the technician level, records were considered part of Y if an analyst reviewed the cluster or batch, but the workunit was not sampled for QA.

Model 1:

- Assumptions: Y and Y' were random. That is, the clusters/batches that were not sampled for a higher review were chosen randomly and the records that were recoded were chosen randomly.
- Inference: The individual change rate estimated the changes present in Y'' and Z.

Model 2:

- Assumptions: Y was random and Y' was not random. That is, the clusters/batches that were not sampled for a higher review were chosen randomly, but the records that were recoded were targeted because they were incorrect.
- Inference: The individual change rate estimated the changes present in Z. There were no remaining defects remaining in Y'' because all of the changes (Y') were corrected.

Model 3:

- Assumptions: Y was not random. That is, the clusters/batches that were not sampled for a higher review were not chosen randomly.
- Inference: The individual change rate estimated the changes present in Y and Z. We know that we corrected Y' changes and removed those cases from our change count.

Model	Y (clusters/batches reviewed, not selected for QA)	Y' (records in clusters/batches not selected for QA, but coded)	Estimation Formula
1	Random	Random	$\frac{p_x*(Y''+Z)}{X+Y+Z}$
2	Random	Not Random	$\frac{p_x*(Z)}{X+Y+Z}$
3	Not Random	n/a	$\frac{[p_x*(Y+Z)]-Y'}{X+Y+Z}$

What are the AFU clerk and technician change rates?

- We calculate the AFU Technician change rates, both individual and overall, and the outgoing quality in the AFU Technician stage, the same way as the BFU Technician change rates only using records from the AFU Technician and AFU Analyst stages.
- We calculate the AFU Clerk change rates, both individual and overall, and the outgoing quality in the AFU Technician stage, the same way as the BFU Clerk change rates, again considering records reviewed by technicians differently than those reviewed by analysts.

Table 20 shows the coding changes in sampled clusters/batches by level of highest review. Records worked in the clerical matching operation only appear in one row, that is, if a clerk and technician both worked the record in BFU in sampled clusters, this record appears only in the

BFU Technician row.

Table 20: Coding Changes in Sampled Clusters/Batches by Level of Highest Review

Stage	Technician Changes	Total records reviewed by Technicians	Analyst Changes	Total records reviewed by Analysts
BFU Clerk	4,859	150,353	1,246	46,315
BFU Technician	n/a	n/a	753	
AFU Clerk	2,060	86,204	657	34,997
AFU Technician	n/a	n/a	847	
Total (n)	6,919	236,557	3,503	81,312

Table 21 shows the overall technician and clerk change rates by stage. Only records to be worked by a clerical user are included; all records that were computer matched and not reviewed are excluded.

Table 21: Overall Change Rate and Outgoing Quality Rate by Stage

Stage	Model 1		Model 2		Model 3	
	Overall Change Rate	Outgoing Quality Rate	Overall Change Rate	Outgoing Quality Rate	Overall Change Rate	Outgoing Quality Rate
BFU Clerk	0.59%	99.41	0.52%	99.48	0.44%	99.56
BFU Technician	0.23%	99.77	0.22%	99.78	0.20%	99.80
BFU Analyst	0.00%	100	0.00%	100	0.00%	100
AFU Clerk	0.95%	99.05	0.11%	99.89	0.30%	99.70
AFU Technician	0.71%	99.29	0.13%	99.87	0.24%	99.76
AFU Analyst	0.00%	100	0.00%	100	0.00%	100

Table 22 shows the number of records which received a final code at the given stage. A record is considered to have a final code at a given stage if it is reviewed as a sampled cluster or batch or the user put a code on the record. Records that were subsampled out or located in surrounding blocks were removed, unless there is a code on that record. Users were not required to review these records.

Table 22: Records Completed by Stage

	Computer Matching	BFU Clerk	BFU Technician	BFU Analyst	AFU Clerk	AFU Technician	AFU Analyst	Total
Records completed	975,094	161,408	96,586	29,471	79,924	84,091	51,442	1,478,016
Percentage of total workload completed	66.0%	10.9%	6.5%	2.0%	5.4%	5.7%	3.5%	100%

Table 23 shows cluster/batch to user ratios for clusters selected for QA. Total includes all clusters reviewed at that stage.

Table 23: Cluster/Batch to User Ratio

Stage	BFU Clusters	AFU Batches
Clerk	47	18
Technician	115	84
Analyst ⁷	113	280

Conclusions

- Matching QA was successful at minimizing errors – The outgoing quality rates are shown above in Table 21. All outgoing quality rates are higher than 99%.
- The workload in the analyst stage was high – The workload in the analyst stage was high for the number of analysts. However, the operation was completed in a reasonable time frame.

Person Followup Quality Assurance

How was person followup quality assurance operationalized?

The quality assurance plan for person followup (PFU) involved two parts: a data edit to ensure completeness of the PFU form and a recontact of the respondent to detect discrepant results.

⁷For this table, an analyst had to work at least 10 clusters/batches in each stage to be considered a working analyst. Several managers worked only a few clusters/batches and are not included in the calculations.

- After the interviewer completed the PFU form, the supervisor (a crew leader) reviewed the PFU form for legibility and completeness. If any part of the form was illegible, then the crew leader contacted the interviewer or respondent to determine the proper answer. If interviewer did not follow any skip pattern or incorrectly left any section of the form blank, then the crew leader contacted the interviewer or sent the questionnaire back to the field for clarification. After the crew leader edit, the crew leader sent the form to the A.C.E. Regional Office (ACERO) where staff reviewed the form for completeness and legibility again.
- To detect discrepant results, the ACE2000 system selected a sample of forms for QA. Three types of cases were sent to PFU QA:
 - A random sample of 1-in-20
 - The first eligible form for each interviewer returned from the field
 - Supervisor-selected cases if an office supervisor suspected discrepant results
- An office QA checker attempted to contact the respondent from the PFU form by telephone. The QA checker asked if a Census Bureau employee recently contacted the respondent. If an interviewer contacted the respondent, then the case passed QA. If an interviewer did not contact the respondent, the QA checker conducted the interview with the respondent.
- If the office QA checker could not contact the respondent by telephone, the ACERO sent the case back to the field to attempt to contact the respondent. The field QA checker attempted to determine if an interviewer had contacted the respondent for the PFU interview using the same procedures as the office QA checker.
- For PFU QA, we limited the respondent recontact to certain cases.
- We only considered completed cases with one respondent eligible for QA recontact.
- We allowed a total of six days to attempt to recontact the respondent: three days by telephone and three days by personal visit.
- Due to time restrictions, we did not allow new field PFU QA cases after 11/17/2000; we did not allow any new telephone PFU QA cases after 11/23.

What are the results from person followup QA?

Table 24 contains the QA selection status of PFU cases. Table 25 contains the workload for PFU QA by A.C.E. Regional Office. Table 26 contains the outcome of cases in PFU QA.

Table 24: QA Selection Status of PFU Cases

QA Status	Randomly Sampled		Targeted	
	Number of Cases	Percentage of PFU QA Cases	Number of Cases	Percentage of PFU QA Cases
Selected	4,491	50.30%	4,438	49.70%
Eligible	4,210	47.15%	4,363	48.86%
Ineligible ⁸	281	3.15%	75	0.84%

Table 25: Workload for PFU QA by A.C.E. Regional Office

A.C.E. Regional Office	Cases in PFU QA (includes ineligible cases)	Total Cases in PFU	Percent PFU Workload in PFU QA	Cases Ineligible for PFU QA	Percent of Ineligible Cases in PFU QA (within ACERO)
Boston	902	6,541	13.79%	68	7.54%
New York	782	7,021	11.14%	38	4.86%
Philadelphia	841	7,502	11.21%	4	0.48%
Detroit	517	4,933	10.48%	15	2.90%
Chicago	537	5,610	9.57%	8	1.49%
Kansas City	384	4,875	7.88%	32	8.33%
Seattle	865	7,697	11.24%	28	3.24%
Charlotte	982	9,174	10.70%	46	4.68%
Atlanta	619	8,397	7.37%	10	1.62%
Dallas	868	10,124	8.57%	76	8.76%
Denver	818	8,733	9.37%	19	2.32%
Los Angeles	814	8,727	9.33%	12	1.47%
Totals	8,929	89,334	10.00%	356	3.99%

⁸We considered a case to be ineligible for PFU QA if there was more than one respondent on the PFU form or if the case was selected after the cutoff date for PFU QA.

Table 26: QA Outcome of Cases in PFU QA

Outcome	Randomly Selected		Targeted	
	Number of Cases	Percentage of Cases in PFU QA	Number of Cases	Percentage of Cases in PFU QA
Pass	3,899	43.67%	4,067	45.55%
Fail ⁹	40	0.45%	84	0.94%
Noninterview	271	3.04%	212	2.37%
Not Eligible	281	3.15%	75	0.84%
Totals	4,491	50.30%	4,438	49.70%

Conclusions

- The A.C.E. PFU QA operation successfully detected discrepant errors.—The QA results in Table 26 show 0.45 percent of all households in the randomly selected sample failed PFU QA. We can assume that the remaining 84,843 cases not randomly selected for QA have the same rate of failure. In addition, we corrected 84 of those cases in the targeted sample.

⁹For PFU QA, a case failed if the respondent had said that the original PFU interviewer did not contact him or her for the original interview. These included discrepant results and those cases that were determined to not be discrepant.

References

- Byrne, R. (May 2000), "MaRCS Specifications for Quality Assurance (QA) for Person Matching", DSSD Census 2000 Procedures and Operations Memorandum Series Chapter S-QA-13.
- Byrne, R. (May 2000), "Quality Assurance for A.C.E. Person Follow-up", DSSD Census 2000 Procedures and Operations Memorandum Series Chapter S-QA-14.
- Byrne, R., Imel, L., Stallone, P. (February 2001), "Accuracy and Coverage Evaluation Survey: Person Interviewing Results", Census 2000 Procedures and Operations Memorandum Series, Chapter B-5*.
- Childers, D. (January 2001), "Accuracy and Coverage Evaluation: The Design Document", Census 2000 Procedures and Operations Memorandum Series, Chapter S-DT-1.
- Davis, M. (July 1991), "Preliminary Final Report for PES Evaluation Project P7: Estimates of P Sample Clerical Matching Error from a Rematching Evaluation", 1990 Coverage Studies and Evaluation Memorandum Series # H-2.
- Davis, M. (July 1991a), "Preliminary Final Report for PES Evaluation Project P10: Measurement of the Census Erroneous Enumerations—Clerical Error made in the Assignment of Enumeration Status", 1990 Coverage Studies and Evaluation Memorandum Series # L-2.
- Haines, D. (May 2000), "Accuracy and Coverage Evaluation Survey: Computer Specification for Person Dual System Estimation (U. S.)", DSSD Census 2000 Procedures and Operations Memorandum Series Chapter Q-29.
- Hogan, H. (September 1993), "The Post-Enumeration Survey: Operations and Results," Journal of the American Statistical Association, pp 1047-1060.
- West, K. (July 1991b), "Final Report for PES Evaluation Project P4: Quality of Reported Census Day Address – Evaluation Follow-up", 1990 Coverage Studies and Evaluation Memorandum Series # D-2.
- "Assessment on Accuracy of Adjusted Versus Unadjusted 1990 Census Base for Use in Intercensal Estimates". Report of the Committee on Adjustment of Postcensal Estimates, United States Bureau of the Census, Washington, D. C., August 7, 1992.
- "Additional Research on Accuracy of Adjusted Versus Unadjusted 1990 Census Base for Use in Intercensal Estimates". Report of the Committee on Adjustment of Postcensal Estimates, United States Bureau of the Census, Washington, D. C., November 25, 1992.

Appendix 1–Percentage of the Weighted E Sample for Each Variable

Table A-1: Percent of the Weighted E Sample by
Census Region

Census Region	Percent of E Sample
Northeast	19.0
Midwest	22.8
South	35.6
West	22.6
Total	100.0

Table A-2: Percent of the Weighted E Sample by Census Regional Office

Census Regional Office	Percent of E Sample
Boston	7.2
New York	6.2
Philadelphia	8.1
Detroit	8.3
Chicago	8.5
Kansas City	7.9
Seattle	7.8
Charlotte	10.5
Atlanta	10.0
Dallas	9.7
Denver	7.1
Los Angeles	8.6
Total	100.0

Table A-3: Percent of the Weighted E Sample by Tenure

Tenure	Percent of E Sample
Owner	67.2
Renter	29.0
Blank	3.8
Total	100.0

Table A-4: Percent of the Weighted E Sample by Type of Enumeration Area

Type of Enumeration Area	Percent of E Sample
Mail Out/Mail Back	82.1
Update/Leave	16.8
List/Enumerate	0.2
Rural Update/Enumerate	0.5
Urban Update/Leave	0.2
Urban Update/Enumerate	0.0
Adds to Address List	0.2
Total	100.0

Table A-5: Percent of the Weighted E Sample by Age and Sex

Age and Sex		Percent of E Sample
Under 18	Male and Female	25.2
18 to 29	Male	7.4
18 to 29	Female	7.5
30 to 49	Male	14.7
30 to 49	Female	15.3
50 and over	Male	12.1
50 and over	Female	15.5
Blank	Male	1.3
Blank	Female	1.2
18 to 29	Blank	0.4
30 to 49	Blank	0.2
50 and over	Blank	0.2
Total		100.0

Table A-6: Percent of the Weighted E Sample by Race Domain

Race Domain	Percent of E Sample
Blank	0.7
American Indians on Reservations	0.2
American Indians Not on Reservations	0.6
Hispanic	15.5
Black	10.3
Pacific Islander	0.2
Asian	3.3
White	67.3
Other Races	0.6
Multiple Races	1.3
Total	100.0

Table A-7: Percent of the Weighted E Sample by Mail Return Rate

Mail Return Rate	Percent of E Sample
Greater than 25 percent	81.6
Less than 25 percent	17.7
List/Enumerate, Rural Update Enumerate, Urban Update/Enumerate	0.7
No Occupied Housing Units or Incomplete Address	0.0
Total	100.0

Table A-8: Percent of the Weighted E Sample by
Percent Mobile Home

Percent Mobile Home	Percent of E Sample
None	76.9
10 Percent or less	9.3
11 to 50 percent	10.4
Greater than 50 percent	3.4
Total	100.0

Table A-9: Percent of the Weighted E Sample by
Percent Multi-Unit

Percent Multi-Unit	Percent of E Sample
None	55.3
10 Percent or less	11.0
11 to 50 percent	14.2
Greater than 50 percent	19.5
Total	100.0

Appendix 2–Equations for Matching Quality Assurance Calculations

Equation 1–BFU Technician Individual Change Rate in Sampled Clusters/Batches

BFU Technician individual change rate, $errT_i = chgsAT_i \div chkAT_i$

where

$chgsAT_i$ are the records worked by technician T_i in a cluster sampled for QA and changed by an analyst, and

$chkAT_i$ are the records worked by technician T_i in a cluster sampled for QA and checked by an analyst.

Equation 2–Adjusted BFU Technician Overall Change Rate for Sampled Clusters/Batches

Adjusted BFU Technician overall change rate, $BFT_err_adj =$

$$\left(\sum_i^n \left(\frac{chgsAT_i - corrCT_i}{chkAT_i} \right) * recsT_i \right) \div \left(\sum_i^n recsT_i \right)$$

where

$chgsAT_i$ is the number of records worked in clusters sampled for QA by technician T_i and changed by an analyst,

$corrCT_i$ is the number of those records reviewed in clusters sampled for QA by technician T_i and changed by an analyst where a clerk coded the record correctly,

$chkAT_i$ is the number of technician T_i 's records in clusters sampled for QA and checked by an analyst, and

$recsT_i$ is the number of records worked in clusters sampled for QA by technician T_i and not reviewed by an analyst.

Equation 3–BFU Clerk Individual Change Rate

$$\text{BFU Clerk individual change rate, } errC_i = \frac{[chgsAC_i + chgsTC_i + (BFT_err_adj * chkTC_i)]}{[chkAC_i + chkTC_i]}$$

where

$chgsAC_i$ are the records worked by clerk C_i and changed by an analyst,
 $chgsTC_i$ are the records worked by clerk C_i and changed by a technician,
 BFT_err_adj is the Adjusted BFU Technician overall change rate,
 $chkTC_i$ is the number of clerk C_i 's records checked only by technicians,
 BFT_err_adj is multiplied by $chkTC_i$ to estimate the potential changes the technicians may have missed in this clerk's work, and
 $chkAC_i$ is the number of records worked by clerk C_i and checked by an analyst.

Equation 4–BFU Technician Overall Change Rate

BFU Technician overall change rate, $BFT_err =$

$$\left(\sum_i^n errT_i * recsT_i \right) \div \left(\sum_i^n recsT_i \right)$$

where

$errT_i$ is the BFU Technician individual change rate for technician T_i , estimated using the change rate for the user in the clusters sampled for QA and
 $recsT_i$ is the number of records worked by technician T_i

Equation 5–Outgoing Quality of the BFU Technician Stage

$$\text{Outgoing quality of the BFU Technician stage} = (1 - BFT_err) * 100$$

Equation 6–BFU Clerk Overall Change Rate

BFU Clerk overall change rate, $BFC_err =$

$$\left(\sum_i^n errC_i * recsC_i \right) \div \left(\sum_i^n recsC_i \right)$$

where

$errC_i$ is the BFU Clerk individual change rate for clerk C_i , and
 $recsC_i$ is the number of records coded by clerk C_i and not reviewed by a technician or an analyst.

Equation 7–Outgoing Quality of the BFU Clerk Stage

$$\text{Outgoing quality of the BFU Clerk stage} = (1 - BFC_{err}) * 100$$